



HAL
open science

Analyse de durées de vie : analyse séquentielle du modèle des risques proportionnels et tests d'homogénéité

Christelle Breuils

► **To cite this version:**

Christelle Breuils. Analyse de durées de vie : analyse séquentielle du modèle des risques proportionnels et tests d'homogénéité. Mathématiques [math]. Université de Technologie de Compiègne, 2003. Français. NNT: . tel-00005524

HAL Id: tel-00005524

<https://theses.hal.science/tel-00005524>

Submitted on 4 Apr 2004

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Université de Technologie de Compiègne
Majeure Contrôle des systèmes

Thèse présentée pour l'obtention du titre de Docteur de l'UTC

Christelle Breuils

ANALYSE DE DURÉES DE VIE :
analyse séquentielle du modèle des risques proportionnels et
tests d'homogénéité.

Le lundi 15 décembre 2003 à 10h30

Mme O. Pons	Rapporteur
M. G. Grégoire	Rapporteur
M. L. Bordes	Directeur de thèse
M. S. Alvarez	
M. J. Antoch	
Mme C. Coccozza	

Remerciements

Loin de moi l'envie de me considérer l'espace de cette page comme une artiste à laquelle on aurait remis un quelconque prix, qui se sentirait obligée de partager son moment de gloire par soucis de modestie publique... Au cours de cette thèse, j'ai sollicité beaucoup de personnes et leur réponse positive à ces multiples demandes nécessite des remerciements, sans eux, le long métrage que constitue un travail de thèse n'aurait pu aboutir.

Dans un premier temps, je remercie les producteurs pour m'avoir accueillie dans leur studio de production, le Laboratoire de Mathématiques Appliquées de Compiègne, au cours de ces trois années de tournage intensif.

Je remercie également M. Grégoire et Me Pons pour avoir accepté d'être rapporteurs de ce mémoire, leur lecture attentive du scénario ainsi que leurs conseils avisés m'ont permis d'améliorer la qualité du scénario aussi bien que la photographie, mais aussi d'envisager de nouvelles voies de développement pour les scénarii futurs.

J'adresse mes remerciements à C. Coccozza, J. Antoch et S. Alvarez pour avoir participé au jury. Je remercie également S. Alvarez pour avoir été mon tuteur pour le monitorat au cours de ces trois années.

Je tiens ensuite à remercier L. Bordes, qui m'a encadré pendant ces trois années, il a su répondre à mes doutes comme à mes attentes tout au long du tournage et a su respecter et même parfois défendre mes divers choix. C'est sans doute grâce à lui que j'ai pu finir dans les délais et je lui en suis reconnaissante.

Je remercie aussi toute l'équipe logistique du Génie Informatique, pour avoir apporté l'aide et les moyens techniques nécessaires à la réalisation de ce long métrage. Je profite de ce paragraphe pour remercier Véronique R. pour avoir préparé les ordinateurs pour la projection finale ; ainsi que Valérie et David pour leur soutien ce jour-là.

Un petit mot ensuite pour les quelques figurants qui furent seulement de passage dans mon bureau, ils m'ont permis de battre le record du nombre de collègues de bureau en trois années ! Ils ont, pour certains, apporté une ambiance chaleureuse et appréciable au sein même de cet espace de travail et pour d'autres, contribué à une ambiance studieuse et reposante. Merci surtout à Christophe et à Pascal donc !

Je tiens également à exprimer toute ma reconnaissance aux quelques acteurs qui ont animé les scènes plus personnelles de cette thèse, j'ai particulièrement apprécié les parties de tarot et de billard, un grand merci donc à Manu, Pierre-Alexandre, David, Valérie et les autres thésards.

Pour les quelques scènes d'action, David, qui s'est improvisé coach sportif, m'a permis de maintenir une bonne physique grâce à la pratique de la natation, du footing et du ping-pong, je l'en remercie, d'autant que sa motivation sportive est bien plus grande que la mienne . . .

Merci à mes parents pour être venus de Tours pour la projection et pour leur soutien. Enfin, merci à Laure pour son soutien et ses nombreuses discussions.

Par ailleurs, je remercie très chaleureusement Manu qui a su me soutenir, m'encourager et m'aider à organiser les avant-premières et la projection finale, la liste de ces aides appréciables n'est évidemment pas exhaustive. Je tenais à ce que les derniers mots de cette page soient pour lui.

Table des matières

I	Analyse séquentielle du modèle de Cox	15
1	Introduction à l'analyse de survie et à l'analyse séquentielle	17
1.1	Les principales fonctions utilisées	19
1.2	Schémas de censure	21
1.3	Le modèle à risques proportionnels de Cox	26
1.4	Procédures séquentielles d'arrêt	32
2	Exemples de règles séquentielles d'arrêt et propriétés	43
2.1	Propriétés	45
2.2	Démonstration de la proposition 2.1.1	46
2.3	Généralisation de la variable d'arrêt	48
3	Théorèmes et propriétés des estimations séquentielles	51
3.1	Convergence presque sûre	53
3.2	Normalité asymptotique	59
4	Analyse séquentielle pour le modèle de Cox	61
4.1	Convergence	63
4.2	Normalité asymptotique : preuve de (C4)	73
5	Normalité asymptotique de la variable d'arrêt N_d	81
5.1	Résultats préliminaires	83
5.2	Hypothèses et représentations de $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$	86
5.3	Normalité asymptotique de $\sqrt{N_d} - \sqrt{n_d}$	91
6	Cadre multidimensionnel	95
6.1	Introductions et hypothèses	97
6.2	Démonstration du théorème	99
6.3	Vérification des conditions (C-I)-(C-IV) pour le modèle de Cox	100
7	Simulations numériques	103
7.1	Simulations et méthodes utilisées	105
7.2	Observations et résultats	105
A	Inégalités maximales	115
A.1	Définition de la norme d'Orlicz	115

A.2	Pseudo-dimension	116
A.3	Inégalités maximales	120
B	Résultats et inégalités diverses	123
B.1	Inégalité portant sur les martingales	123
B.2	Rappels de résultats	124
C	Condition d'Anscombe	125
C.1	Continuité uniforme en probabilité et première condition d'Anscombe	125
C.2	Papiers postérieurs	126
II	Tests d'homogénéité	129
8	Cadre de l'étude	131
8.1	Modèle adopté et propriétés	133
8.2	Test d'homogénéité	135
9	Comportement asymptotique de la statistique de test	143
9.1	Définitions, notations et asymptotique de \hat{T}_n	145
9.2	Démonstration du théorème 9.1.1	146
D	Quelques rappels de résultats	159
	Conclusions et perspectives	161
	Bibliographie	165

Notations

Pour N , un ensemble, on note :

- $\sigma\{N\}$ la tribu engendrée par N ;
- $\mathbb{1}_N$ la fonction indicatrice de l'ensemble N ;
- N^c le complémentaire de N dans l'espace sous-jacent ;
- $\text{Card}(N)$ le nombre d'éléments de l'ensemble N , s'il est fini.

Si (Ω, \mathcal{T}) est un espace probabilisable, on note :

- \mathbb{P} la mesure de probabilité qui est attachée ;
- \mathbb{E} l'espérance mathématique associée à \mathbb{P} .

Convergences

On note :

- $\xrightarrow{\mathcal{D}}$ la convergence en loi ;
- $\xrightarrow{\mathcal{P}}$ la convergence en probabilité ;
- $\xrightarrow{p.s.}$ la convergence presque sûre ;
- \xrightarrow{c} la convergence complète.

Lois usuelles

On note :

- χ_d^2 la loi du chi-deux à d degrés de liberté ;
- $\mathcal{N}(\mu, \sigma^2)$ la loi normale de moyenne μ et de variance σ^2 ;
- $\mathcal{W}(a, b)$ la loi de Weibull de paramètres a et b ,
- $\mathcal{U}([a, b])$ la loi uniforme sur $[a, b]$,
- τ_p la loi de Student à p degrés de libertés.

Pour a et b deux éléments d'un espace totalement ordonné, on note :

$a \wedge b$ le maximum de a et b ;

$a \vee b$ le minimum de a et b .

On note :

\mathcal{O}, o le « grand o » et « petit o » de Landau ;

\mathcal{O}_P, o_P le « grand o » et « petit o » pour la convergence en probabilité ;

L^p l'espace des fonctions de puissance p intégrables ($1 \leq p \leq +\infty$) ;

ℓ^p l'espace des suites de puissance p sommables ($1 \leq p \leq +\infty$) ;

$\text{Leb}(A)$ la mesure de Lebesgue de l'élément A .

Pour f une fonction de A à valeurs dans B et t un élément de A , on note :

$f(t^-)$, la limite à gauche en t de f ,

$C^2(B)$, l'ensemble des fonctions de classe C^2 sur B .

Pour E un espace normé, $a \in E$ et $\eta > 0$, on définit :

$B(a, \eta)$ la boule de centre a et de rayon η ,

$\partial B(a, \eta)$ la frontière de $B(a, \eta)$.

Notations matricielles

On note :

$\mathcal{M}_p(\mathbb{R})$ l'ensemble des matrices réelles carrées de dimension p ,

M^T la matrice transposée de la matrice M ,

$\|\cdot\|$ une norme sur \mathbb{R}^p ou $\mathcal{M}_p(\mathbb{R})$ (selon le contexte).

D'après le récapitulatif qu'il dresse dans [Gho91], Ghosh date les rudiments de l'analyse séquentielle aux travaux de Huyghens, Bernoulli, De Moivre et Laplace sur le problème de la ruine du joueur. Les concepts de procédures d'inspection séquentielles d'échantillonnage sont apportés ensuite par Dodge et Romig [DR29], les diagrammes de contrôle de qualité introduits par Shewhart [She31] et le modèle en deux étapes (*two-stage design*) par Thompson [Tho33] en 1933. Ghosh décrit ensuite la période 1943-1950 comme la naissance et l'enfance du domaine, au cours desquelles Wald et Barnard [Wal46, Bar46] ont chacun introduit séparément les tests à rapports de probabilité séquentiels (*sequential probability ratio test*), Wald et Wolfowitz [WW48] montrèrent leur optimalité et Haldane et Stein [Hal45, Ste45] montrèrent que les méthodes séquentielles peuvent être utilisées pour résoudre certains problèmes d'estimation ponctuelles ou par intervalles (voir historiques concernant l'analyse séquentielle par exemple dans [Lai01, Sie03]).

C'est ainsi qu'en 1945 dans le cadre de l'estimation de la moyenne, Stein [Ste45] propose une procédure en deux étapes visant à fournir un intervalle de confiance de longueur précisée par le statisticien, dans le cas d'un échantillon issu d'une loi gaussienne. La longueur de l'intervalle de confiance obtenu ne dépend pas de la variance des observations mais uniquement de la variance empirique des observations initiales.

Stein [Ste49] et Anscombe [Ans53] suggèrent alors des méthodes entièrement séquentielles où l'on calcule la variance après chaque nouvelle observation, et où la procédure est stoppée lorsque la taille empirique de l'intervalle de confiance est plus petite qu'une longueur désirée (pour des tailles supérieures à un seuil initial). Les premières propriétés concernant ces règles d'arrêt sont alors étudiées. Stein et Anscombe montrent de façon heuristique que l'intervalle de confiance alors considéré a le taux asymptotique de couverture souhaité et que la procédure d'arrêt possède une propriété dite d'efficacité asymptotique lorsque la longueur souhaitée tend vers 0, dans le sens où son espérance est asymptotiquement équivalente à ce qui serait requis si l'on connaissait la variance et qu'on l'utilisait pour calculer la taille de l'échantillon requise pour obtenir un tel niveau de précision.

Chow et Robbins [CR65] apportent ensuite un cadre formel aux démonstrations et ouvrent la porte à de nombreuses extensions et généralisations. La période de 1951 à 1990 est décrite par Ghosh comme étant la période joignant l'adolescence à l'âge adulte et est marquée par des contributions importantes et de grands développements dans divers sujets. Citons parmi elles un article d'Anscombe [Ans52] qui vise à établir un théorème de la limite centrale adapté à ce type de processus. Il est

possible d'explorer une partie de la large diversité des règles d'arrêt par la lecture de [GMS97], qui présente les méthodes séquentielles et démontre leurs propriétés. Dmitrienko et Govindarajulu [DG00] fournissent un outil séquentiel complet dans le cas de maximum de vraisemblance (paramétrique) et démontrent les propriétés de certains estimateurs.

Parallèlement à cette avancée de l'analyse séquentielle, dans le milieu du siècle dernier, alors que l'analyse de durée de vie n'existait principalement que par la contribution de Greenwood [Gre26], on voit apparaître de nouveaux modèles et développements comme ceux de Weibull [Wei51], de Kaplan et Meier [KM58], ainsi que le modèle de Cox [Cox72], qui cotoie le développement de la théorie des martingales en temps continu. Il sera largement utilisé dans le domaine médical et en analyse de durée de vie en général.

Cox [Cox72] donne un estimateur avec le premier article présentant le modèle, et justifie ensuite son calcul à l'aide du concept de vraisemblance partielle [Cox75]. Il montre notamment certaines propriétés de convergence des estimateurs du maximum de vraisemblance partielle, sous des conditions restrictives. De nombreux auteurs lui succèdent pour fournir des discussions sur les théorèmes de la limite centrale en analyse de durée de vie [Gil80, Tsi81, Nae82]. Andersen et Gill [AG82] apportent une étude basée sur la théorie des martingales, car ils considèrent qu'elle émerge tout naturellement de la formulation du modèle.

Notons dès à présent que le modèle de Cox permet de prendre en compte une certaine hétérogénéité des données, puisque chaque individu est caractérisé par un vecteur de variables explicatives. Lorsque l'on dispose de données pour lesquelles on souhaite appliquer le modèle de Cox, un test préliminaire consiste souvent à vérifier que ces données sont effectivement hétérogènes par le biais du test de Wald [KM97], par exemple.

Sellke et Siegmund [SS83] présentent un article qui relie les domaines du séquentiel et de l'analyse de durée de vie. Ils considèrent ici le modèle de Cox à entrées retardées et donnent des résultats de convergence asymptotique des estimateurs à l'aide de techniques utilisant des martingales. Cette théorie sera généralisée dans [BGY97].

À notre connaissance aucun résultat concernant l'estimation séquentielle au sens de [GMS97] pour le modèle de Cox n'est paru. Notre première préoccupation consiste alors à rassembler ces deux axes que sont l'analyse de survie et le modèle de Cox ainsi que l'analyse séquentielle, plus particulièrement l'estimation séquentielle.

Ainsi, nous rappellerons succinctement dans le chapitre 1 les divers outils de l'analyse de survie et les fonctions principalement étudiées, lorsque l'on souhaite étudier le temps de survenue d'un événement d'intérêt \mathcal{E} . La plupart du temps, il se révèle impossible d'observer directement \mathcal{E} , nous définissons alors les phénomènes de censure et donnons un cadre formel au cas particulier de la censure à droite. Les données et leurs caractéristiques principales étant définies, nous aborderons ainsi l'étude du modèle de Cox ainsi que le rappel des fonctions et estimateurs qui peuvent

intervenir. Les procédures séquentielles sont alors présentées, nous verrons quelles propriétés il est possible d'attacher aux variables d'arrêt.

Nous définirons alors une classe de variables d'arrêt, notée \mathcal{A} . Nous étudierons au cours du second chapitre un élément particulier de \mathcal{A} qui sera implémenté dans les simulations numériques.

Pour des règles d'arrêt de \mathcal{A} , il est possible de généraliser le théorème de [DG00] qui nous fournit des résultats de convergence en distribution, l'étude de cette généralisation fait l'objet du chapitre 3, nous nous plaçons dans un cadre semi-paramétrique et unidimensionnel. Nous dressons alors un ensemble de conditions que le modèle doit vérifier. Une partie de ce chapitre est consacrée à la convergence complète qui se révèle un outil clé de l'analyse séquentielle.

Vient ensuite dans le chapitre 4 l'application des résultats du chapitre 3 au modèle de Cox. Nous verrons que, sous des hypothèses analogues à celles de [AG82], la démonstration des conditions données au chapitre 3 nécessite des démonstrations pointues. Ainsi, l'une d'elles utilise des inégalités maximales inspirées de Pollard [Pol90], ou certaines inégalités résultant de propriétés des martingales. La seconde fait intervenir une condition de type Anscombe, pour laquelle nous utiliserons principalement le théorème de Rebolledo, l'inégalité de Kolmogorov et une inégalité exponentielle pour les martingales. Là encore, nous devons faire un lien entre la convergence complète et la continuité uniforme en probabilité.

Il peut se révéler intéressant de connaître la variabilité que subit l'estimateur de la taille d'échantillon N_d (d représente l'amplitude de l'intervalle de confiance souhaité) par rapport à sa valeur idéale n_d (*i.e.* qui est la taille de l'échantillon pour une variance du paramètre de régression connue), ce qui revient à étudier le comportement asymptotique de la variable aléatoire donnée par $\sqrt{N_d} - \sqrt{n_d}$. Nous verrons, par un résultat issu de [GMS97], qu'un tel calcul se ramène à l'étude du comportement asymptotique de l'estimateur de la variance asymptotique. Après quelques résultats préliminaires techniques, nous donnons deux représentations du processus d'intérêt. La première est prédestinée au calcul de la loi asymptotique. La seconde nous aidera à montrer une certaine robustesse du comportement asymptotique de l'estimateur de la variance asymptotique, lorsque celui-ci est indexé par une variable aléatoire donnée par une règle d'arrêt élément de \mathcal{A} . Nous montrerons alors le principal résultat de ce chapitre, c'est-à-dire la normalité asymptotique de $\sqrt{N_d} - \sqrt{n_d}$, dans le cadre d'observations issues de variables indépendantes et identiquement distribuées.

Un court chapitre suit avec le but d'étendre les résultats unidimensionnels au cas multidimensionnel ; nous montrons que grâce à certaines propriétés du déterminant et aux démonstrations effectuées en dimension un, nous ne rencontrons pas de difficultés majeures pour adapter nos démonstrations à ce nouveau cadre.

Enfin, un chapitre présentant des simulations numériques clôt cette première partie et permet ainsi de vérifier certains résultats établis.

Notons de plus que des annexes viennent éclairer l'utilisation de certains outils mathématiques. L'annexe A décrit brièvement le moyen d'établir des inégalités exponentielles pour des processus empiriques *via* le contrôle de l'entropie. L'annexe B est

essentiellement consacrée à une inégalité exponentielle pour les martingales, tandis que l'annexe C est une discussion sur le concept de continuité uniforme en probabilité.

Dans une seconde partie, nous nous intéressons à un test d'homogénéité basé sur un modèle de régression non paramétrique décrit par McKeague et Utikal [MU90].

Dans leur article concernant les tests d'homogénéité basés sur des processus de Poisson, Ng et Cook [NC99] dressent un historique de l'étude du type de données nécessitant de tels tests que nous reprenons succinctement ici.

Dans les études pour lesquelles les pannes ne sont pas nécessairement fatales, on voit souvent apparaître des données dites d'événements récurrents [KP80]. De nombreux domaines d'applications, comme l'épidémiologie [GSB80], l'économie [Lan90] et la fiabilité [KLR91] sont ainsi concernés. Les exemples les plus couramment cités en épidémiologie traitent des crises d'asthme, des crises d'épilepsie, des admissions dans les hôpitaux, ou encore des apparitions de tumeurs cancéreuses.

Le point qui relie ces données réside dans la présence d'hétérogénéité non observée entre les sujets (par exemple, une variation biologique non expliquée par les variables explicatives dont on dispose [Aal94]). C'est ainsi que se développèrent de nombreuses études afin d'analyser ces données. On distingue essentiellement deux approches : l'une est basée sur les modèles à effets aléatoires (de fragilité [Aal88, CC85, Kle92, NC97, NGAS92]) et la seconde sur les modèles marginaux [LN95]. Lorsqu'elle tient compte de l'hétérogénéité, la formulation du modèle est plus complexe et il s'ensuit généralement une baisse d'efficacité des méthodes d'estimation. À l'inverse, estimer un modèle de données homogènes *via* des données hétérogènes peut produire un biais important dans les estimations des coefficients de régression [BHC88, SK86]. C'est ainsi que tester le caractère homogène des données peut se révéler une étape importante qui mène à une spécification du modèle plus appropriée.

Les statistiques basées sur les fonctions score sont fréquemment utilisées pour construire de tels tests, elles possèdent effectivement des propriétés intéressantes et nécessitent des hypothèses minimales sur la distribution de l'effet aléatoire [CA95, Cox83, Dea92, DL89, Gra95, JGC95, Lia87]. En outre, dans la plupart des cas, les tests sont aisés à implémenter.

Nous considérons ici un modèle général où la fonction de risque, conditionnellement à une covariable Z , est une fonction $\alpha(\cdot, Z)$ pour laquelle la dépendance à Z n'est pas spécifiée par une fonction de lien, contrairement au modèle de Cox.

L'hétérogénéité est alors modélisée par une variable de fragilité ε (non observable) qui vient modifier la fonction de risque d'un individu ayant la covariable Z en $\varepsilon\alpha(\cdot, Z)$, augmentant ou diminuant ainsi son risque.

Le test d'homogénéité consiste alors à vérifier que toutes les variables de fragilité sont égales à 1. La structure de corrélation des variables de fragilité est spécifiée par une matrice W connue. Ainsi pour tester un effet « famille » (voir [CJG97]), on impose au terme w_{ij} de W d'être égal à 1 si les individus i et j sont issus de la même famille et 0 sinon. Dans cette étude, nous nous sommes limités à de tels effets, appelés « effets de groupes ».

Dans une première partie, nous rappelons brièvement le modèle étudié ainsi que ses propriétés, qui ont été étudiées par McKeague et Utikal [MU90]. Nous verrons ainsi les hypothèses nécessaires au bon comportement de l'estimateur de Beran [Ber81]. Puis, nous donnerons notre cadre de travail, en définissant la matrice W et les propriétés requises. À partir du modèle ainsi établi, nous construisons un test d'homogénéité basé sur la fonction score, puis nous proposons une statistique de test.

Le but du second chapitre consiste principalement à démontrer le théorème de normalité asymptotique de la statistique de test qui est présenté après quelques notations et remarques. Les techniques utilisées sont principalement basées sur des propriétés de martingales des processus mis en jeu.

Première partie

Analyse séquentielle du modèle de Cox

Chapitre 1

Introduction à l'analyse de survie et à l'analyse séquentielle

Sommaire

1.1	Les principales fonctions utilisées	19
1.2	Schémas de censure	21
1.2.1	Censure à droite	22
1.2.2	Censure à gauche	26
1.2.3	Censure par intervalles	26
1.3	Le modèle à risques proportionnels de Cox	26
1.3.1	Description du modèle et introduction	26
1.3.2	Estimateurs	27
1.4	Procédures séquentielles d'arrêt	32
1.4.1	Préoccupations menant à diverses procédures d'arrêt . . .	33
1.4.2	Définitions générales	36
1.4.3	Critères d'arrêt dans le cadre semi-paramétrique	39

Dans beaucoup de domaines tels que la biologie, la santé publique, l'épidémiologie ou même l'industrie, modéliser le temps de survenue d'un événement apporte une information utile et appréciable. Les outils statistiques ainsi utilisés ont généré deux visions de ce problème dues à des domaines d'application distincts que sont l'analyse de durées de vie et la fiabilité ; les concepts utilisés sont semblables, on distingue cependant des variations dans le vocabulaire d'usage. Au cours de ce mémoire, nous adopterons préférentiellement le point de vue de l'analyse de survie. Il correspond effectivement à celui choisi par Cox lorsqu'il introduisit le modèle à risques proportionnels en 1972. Nous nous plaçons dans un cadre général où l'on souhaite étudier la durée T écoulée entre le début de l'étude et la réalisation d'un événement d'intérêt \mathcal{E} . C'est alors la nature de \mathcal{E} qui distingue l'analyse de survie (il s'agira alors plutôt de mort, d'apparition de tumeur, de développement ou occurrence d'une maladie, de conception, ...) de la fiabilité (il sera question d'étude d'événements relatifs au domaine industriel comme les pannes, l'apparition de symptômes qui présagent d'une défaillance, ...).

Nous définissons dans un premier temps les fonctions les plus courantes dans le domaine de l'analyse de survie. Il se révèle parfois impossible d'observer avec exactitude la variable T , appelée durée de vie. Ce peut être le cas par exemple lors d'études cliniques menées sur l'effet d'un médicament : lorsque les prélèvements de sang sont effectués à intervalles fixes, on ne peut alors renseigner de manière précise la date de l'éventuel changement de situation et la seule information apportée pour la durée de vie est un intervalle de temps. C'est ainsi que les schémas de censure permettent de tenir compte de l'information exacte fournie par cette étude ; nous verrons qu'il existe plusieurs situations se reflétant dans le modèle statistique par plusieurs types de censure. L'étude d'un modèle soumis à un schéma de censure nécessite quelques adaptations au niveau du calcul des estimateurs, elles feront l'objet d'un court résumé dans le cas du modèle de Cox. Nous définirons enfin les procédures séquentielles et décrirons leurs principales propriétés.

Dans le but de fixer les notations, ce mémoire débute par un rappel des différentes fonctions utilisées en analyse de survie, il ne prétend aucunement à une quelconque exhaustivité et son plan est essentiellement basé sur celui de [KM97], où l'on peut également trouver des exemples de données réelles commentés. Par ailleurs, les notations et conventions utilisées pour les paramètres des différentes lois sont calquées sur celles de Matlab, afin de faciliter la cohérence avec la partie numérique présentée au cours du chapitre 7.

1.1 Les principales fonctions utilisées

On désigne par T , le temps écoulé jusqu'à l'apparition de \mathcal{E} . Dans toute cette partie, on se place sur $(\Omega, \mathcal{T}, \mathbb{P})$ un espace probabilisé. T se modélise alors par une variable aléatoire réelle prenant uniquement des valeurs positives ; on supposera en outre que la variable T est continue. La première fonction étudiée est la fonction de survie ; elle désigne, pour un instant t fixé, la probabilité que l'individu étudié vive après t (*i.e.* la probabilité que \mathcal{E} n'ait pas lieu au cours de l'intervalle de temps $[0, t]$). Dès lors et dans toute la suite, on supposera que l'étude débute à l'instant 0 (il est aisé de s'y ramener à l'aide d'une translation).

Définition 1.1.1. *La fonction de survie S , relative à T , variable aléatoire de durée de vie se définit par :*

$$S : \begin{cases} \mathbb{R}_+ & \longrightarrow [0, 1] \\ t & \longmapsto \mathbb{P}(T > t). \end{cases}$$

Remarque 1.1.1. *S s'exprime simplement en fonction de F , la fonction de répartition de T*

$$\begin{aligned} S(t) &= \mathbb{P}(T > t) \\ &= 1 - F(t). \end{aligned}$$

Par définition, S est décroissante, continue à droite et de limites 0 et 1 respectivement en $+\infty$ et 0.

Précisons dès à présent que dans la pratique, l'étude ne peut qu'être menée sur un intervalle de temps fini du type $[0, \tau]$ où τ désigne ainsi la durée de l'étude. Nous préciserons plus avant les conditions que doit vérifier τ afin d'assurer une certaine cohérence du modèle.

Les distributions paramétriques les plus couramment utilisées pour T sont les lois de Weibull et exponentielle. La fonction de densité de Weibull de paramètres a et b , est donnée par

$$f_{(a,b)} : x \mapsto abx^{b-1}e^{-ax^b} \mathbb{1}_{\mathbb{R}_+}(x),$$

où $\mathbb{1}_A$ désigne la fonction indicatrice de l'ensemble A . Il est alors possible d'exprimer la loi exponentielle de paramètre μ comme un cas particulier de la loi de Weibull, en prenant $b = 1$ et $a = \mu$.

La loi de Weibull sera en outre choisie pour modéliser les phénomènes de censure dans le cadre des simulations ; cela tient essentiellement à la flexibilité de la fonction de survie obtenue en fonction des paramètres.

Définition 1.1.2. *On définit la fonction de risque par*

$$\forall t \geq 0, \lambda(t) = \lim_{h \rightarrow 0^+} \frac{\mathbb{P}(t \leq T < t + h | T \geq t)}{h}.$$

Si T admet une densité f (par rapport à la mesure de Lebesgue), on a $\lambda = f/S$. Comme son nom le laisse présager, une vision intuitive de la fonction de risque λ prise en t se résume (de façon informelle) à l'intensité d'apparition de \mathcal{E} à l'instant t (sachant que \mathcal{E} n'a pas eu lieu au cours de la période $[0, t]$).

Remarque 1.1.2. *On suppose désormais que T est absolument continue par rapport à la mesure de Lebesgue. Les définitions qui suivent seront données dans ce cadre restrictif, pour plus de concision.*

Soit donc T une variable aléatoire positive de densité f . Avec les mêmes notations que précédemment, on peut écrire, de façon informelle, pour $t \geq 0$, et dt « petit »,

$$\lambda(t)dt = \mathbb{P}(t \leq T < t + dt | T \geq t) + o(dt), \quad (1.1)$$

ce qui permet une interprétation en terme de probabilité pour la quantité $\lambda(t)dt$.

C'est ainsi que l'allure de la fonction de risque apporte une information immédiate sur les caractéristiques de l'objet étudié ; elle indique en effet comment les chances de survenue de \mathcal{E} varient au cours de sa vie. On pourra ainsi déduire d'un modèle adopté pour un composant industriel qu'il est assujéti à une forme de vieillissement si la courbe représentative de λ est croissante. Les courbes décroissantes peuvent par exemple souligner une faible probabilité d'apparition de \mathcal{E} exprimant des risques de rejet suite à une transplantation dans le cas d'une étude clinique. Le plus fréquemment en fiabilité, ce sont les courbes dites en baignoire qui caractérisent la vie d'un composant : la première partie représente les défauts de jeunesse, la seconde, la période de vie utile, et la dernière, le vieillissement. Elles sont aussi utilisées pour modéliser la durée de vie d'un individu suite à une opération (la partie croissante correspondant aux risques d'infections ou de rejets suite à l'intervention). Grâce à ces propriétés, il apparaît que l'allure de la fonction de risque est plus propice à une étude qualitative que ne l'est le graphe de la fonction de survie.

Définition 1.1.3. *La fonction de risque cumulé est définie sur \mathbb{R}_+ par :*

$$\Lambda : x \mapsto \int_0^x \lambda(t)dt.$$

Remarque 1.1.3. *Une distribution est déterminée de manière équivalente par sa densité, sa fonction de répartition, sa fonction de risque ou encore sa fonction de risque cumulé (lorsqu'elles existent). Il est effectivement possible de se ramener à l'une ou l'autre de ces fonctions à l'aide d'opérations simples.*

En effet, on peut établir les relations suivantes entre les différentes fonctions précédemment définies (sous réserve de leur existence). Ainsi, pour $t \geq 0$, on a :

$$\begin{aligned}\lambda(t) &= \frac{f(t)}{S(t)} = -\frac{d}{dt} \log S(t), \\ \Lambda(t) &= \int_0^t \lambda(u) du = -\log S(t), \\ S(t) &= \exp(-\Lambda(t)).\end{aligned}$$

Comme précédemment notifié, il est en pratique parfois impossible d'observer l'événement \mathcal{E} de référence dès son apparition, et donc les réalisations de la variable T . L'introduction d'une nouvelle variable dite de censure permet alors de modéliser l'information effective. Différents cas se présentent, ils correspondent à divers types de censure.

1.2 Schémas de censure

On trouve principalement deux phénomènes dont les effets se traduisent par des données incomplètes ou observées partiellement dans les applications en fiabilité et en durée de survie [Hub00]. Elles correspondent effectivement à deux situations distinctes : la censure et la troncature. Afin d'explicitier brièvement ces notions, reprenons un exemple de [ABGK93, p.152]. Considérons un jeu de données représentant la durée de survie parmi un groupe constitué d'insulino-dépendants au Danemark. On peut y trouver, distingués par tranche d'âges et par sexe, les effectifs des personnes diabétiques vivantes au 1^{er} juillet 1973 ainsi que les effectifs des décès durant la période juillet 1973 - décembre 1981. Supposons que l'on souhaite étudier la mortalité comme fonction de la durée de l'insulino-dépendance. Il est alors impossible de savoir, pour un patient donné, la date exacte du début de la maladie, l'unique information apportée est son antériorité au début de l'étude. Les distributions qu'il est judicieux d'étudier ici nous sont données par la loi de X conditionnellement à $X > V$, où V représente le temps écoulé entre l'apparition de la maladie et le début de l'étude. En effet, il se peut qu'il y ait des patients diabétiques non considérés dans l'étude puisque décédés à son début, mais pour lesquels le début de la maladie est postérieur à celui de certaines personnes incluses dans l'étude. Le traitement des données sans tenir compte de ce fait sera biaisé puisqu'il favorise ici les temps de maladie longs. Ces données sont tronquées à gauche car les temps de maladie inférieurs à V ne sont pas observables. Pour la distinguer de la censure, prenons l'exemple (toujours issu de [ABGK93, p.24]) d'une étude sur les babouins au Kenya. Ces animaux présentent la particularité de dormir dans les arbres et mettent pied à terre dans la journée (fait répertorié uniquement si la moitié au moins des membres du groupe est descendue), événement duquel on souhaite répertorier la durée ainsi que les dates. Il se peut que l'observateur en charge de la mission arrive une fois que la communauté est déjà descendue, il ne peut alors qu'affirmer que la descente a eu lieu avant une certaine date. Il s'agit ici de censure à gauche.

1.2.1 Censure à droite

Supposons que l'on souhaite observer l'effet d'un traitement donné sur une population donnée d'individus. Il se peut que certains abandonnent l'étude. Radier ces individus des données serait biaiser le traitement statistique, la censure dite à *droite* permet de modéliser l'information exacte apportée. Il s'agit d'attribuer une valeur c à la durée jusqu'à l'abandon de l'étude par l'individu considéré, qui sera modélisée par la réalisation d'une variable aléatoire C , dite de censure. La valeur alors observée pour l'individu démissionnaire nous est simplement donnée par $x = c \wedge t$, où t représente la réalisation de la durée de vie. De surcroît, x n'est pas la seule information apportée ; en effet, il est en général possible de savoir si l'on a observé \mathcal{E} , c'est-à-dire de connaître la valeur de $\delta = \mathbb{1}_{\{x \leq c\}}$, qui indique s'il y a eu censure. Il existe cependant des études où l'information sur δ peut manquer [MS98].

Modèle statistique 1.

Soit un n -échantillon d'individus soumis à une censure à droite. Les observations sont alors constituées des variables aléatoires $(X_i, \delta_i)_{1 \leq i \leq n}$, où

δ_i est la fonction indicatrice de l'observation réelle de l'événement d'intérêt,

T_i et C_i sont respectivement les variables aléatoires de durée de vie et de censure,

$X_i = C_i \wedge T_i$ est la durée effectivement observée.

Dans ce cas, les observations sont les $(X_i, \delta_i)_{1 \leq i \leq n}$, et l'on sait que

$$\begin{aligned} X_i &= T_i & \text{si } \delta_i &= 1, \\ X_i &< T_i & \text{si } \delta_i &= 0. \end{aligned}$$

Il est impossible d'effectuer un traitement statistique sans hypothèse sur les censures [ABGK93, p.49]. Prenons alors pour acquis le fait que pour tout instant t , l'expérience des données de survie n'est en aucun cas affectée par la censure, ou même, les expériences passées. Pour ce faire, il est nécessaire d'apporter une formalisation à la notion de passé, *via* les filtrations.

Définition 1.2.1.

Soit \mathcal{F} une tribu. On dit que la famille de sous-tribus $\{\mathcal{F}_t; t \geq 0\}$ est une filtration si :

$$s \leq t \Rightarrow \mathcal{F}_s \subset \mathcal{F}_t.$$

On définit également \mathcal{F}_{t-} comme étant la tribu engendrée par $\cup_{h>0} \mathcal{F}_{t-h}$.

Il apparaît alors naturel de considérer l'information apportée au cours du temps comme une filtration, ce qui permet l'utilisation des techniques mathématiques s'y rapportant et de décrire le fait que l'information passée est contenue dans l'information actuelle. On définit également la notion de processus prévisible.

Définition 1.2.2. Soit $(\Omega, \mathcal{F}, \mathbb{P})$, un espace probabilisé muni d'une filtration $\mathcal{F} = \{\mathcal{F}_t; t \geq 0\}$. On définit la tribu sur $\mathbb{R}_+ \times \Omega$ engendrée par tous les ensembles de la forme

- $[0] \times A; A \in \mathcal{F}_0$,
- et $]a, b] \times A, 0 \leq a < b < \infty, A \in \mathcal{F}_a$,

comme étant la tribu prévisible engendrée par la filtration $\{\mathcal{F}_t; t \geq 0\}$.

Un processus X est dit prévisible par rapport à une filtration \mathcal{F} , si, en tant qu'application de $\mathbb{R}_+ \times \Omega$ dans \mathbb{R} , il est mesurable par rapport à la tribu prévisible engendrée par \mathcal{F} .

Enfin, on dit du processus X qu'il est adapté à la filtration \mathcal{F} , si, pour $t \geq 0$, $X(t)$ est \mathcal{F}_t -mesurable.

Pour un individu i donné, on définit, pour $t \geq 0$, les variables suivantes

$$\begin{aligned} Y_i(t) &= \mathbb{1}_{\{X_i \geq t\}}, \\ N_i(t) &= \mathbb{1}_{\{X_i \leq t, \delta_i = 1\}}. \end{aligned}$$

$Y_i(t)$ indique si i est un individu à risque à l'instant t (*i.e.* s'il n'a pas subi \mathcal{E} durant la période $[0, t]$) et $N_i(t)$ indique si l'on a observé \mathcal{E} avant l'instant t . On pose, pour n entier et t positif, $\bar{N}(t) = \sum_{i=1}^n N_i(t)$.

$\bar{N}(t)$ détermine alors le nombre d'événements \mathcal{E} observés au cours de l'intervalle $[0, t]$. La notion de processus de comptage et la théorie qui lui est associée (voir par exemple [FH90]) permettent d'obtenir des résultats sur les processus dans lesquels ils interviennent. Ils sont fréquemment utilisés en analyse de survie, et définis comme suit.

Définition 1.2.3. Un processus de comptage est un processus stochastique $\{N(t); t \geq 0\}$ adapté à une filtration $\{\mathcal{F}_t; t \geq 0\}$, avec $N(0) = 0$, $N(t)$ fini presque sûrement (pour t positif), et pour lequel les trajectoires sont continues à droite, constantes par morceaux, et possèdent des discontinuités sous forme de sauts de taille un.

Propriété 1.2.1. Pour $1 \leq i \leq n$, N_i est un processus de comptage adapté à la filtration naturelle associée $\mathcal{F}_t = \sigma\{N_i(u); 0 \leq u \leq t, 1 \leq i \leq n\}$.

Démonstration :

Par définition, le processus est adapté à la filtration. Pour i entier naturel, on a $N_i(0) = 0$, puisque X_i est une variable à densité et donc diffuse (ce qui implique en particulier que $\mathbb{P}(X_i = 0) = 0$). N_i est fini presque sûrement, puisque plus petite que 1. Enfin, N_i est continue à droite, constante par morceaux et a au plus un point de discontinuité, de taille 1. \square

On suppose généralement [ABGK93, KM97, AG82] que le modèle vérifie les propriétés suffisantes pour que l'heuristique suivante soit vérifiée

$$\mathbb{E}(dN_i(t)|\mathcal{F}_{t-}) = Y_i(t)\lambda_i(t)dt. \quad (1.2)$$

On se propose dès lors d'expliciter cette heuristique à l'aide d'arguments que l'on justifie ici par le biais d'exemples issus de [FH90]. Cela nous permettra en outre d'introduire les notions de martingales et de compensateur. On conserve le cadre précédemment défini, et l'on se place dans le cas particulier où les variables aléatoires T et C sont indépendantes. Pour dt « petit », on a

$$\mathbb{P}(t \leq T < t + dt | T \geq t, C \geq t) = \lambda(t)dt + o(dt),$$

ce n'est autre qu'une adaptation de (1.1) au cas censuré. Si l'on note $N(t^-)$, la limite à gauche en t de N , on obtient

$$\lambda(t)dt \approx \mathbb{E}(N((t+dt)^-) - N(t^-) | T \geq t, C \geq t),$$

puisqu'entre autres, la quantité $N((t+dt)^-) - N(t^-)$ prend uniquement les valeurs 0 et 1. Si l'on considère le processus A défini pour $t \geq 0$, par

$$A(t) = \int_0^t \mathbb{1}_{\{X \geq u\}} \lambda(u) du,$$

où $X = T \wedge C$, on peut déduire, d'après la formule qui précède qu'il devrait approcher le nombre de sauts de N au cours de $]0, t]$. Il est en effet possible d'établir que $\mathbb{E}N(t) = \mathbb{E}A(t)$ (voir [FH90, p.19]). On note alors que le processus M , défini au temps t par $M(t) = N(t) - A(t)$ est d'espérance nulle. En outre, la moyenne conditionnelle de $N(t+s) - N(t)$ sachant l'information contenue dans $N(u)$ et $\mathbb{1}_{\{X \leq u, \delta=0\}}$ (pour u élément de $[0, t]$), correspond à la moyenne conditionnelle de $A(t+s) - A(t)$, ce qui nous permet de déduire l'heuristique annoncée. On remarque en effet que $\mathbb{E}(dN(s)|\mathcal{F}_{s-}) = dA(s)$, où $\mathcal{F}_{s-} = \sigma\{N(u), \mathbb{1}_{\{X \leq u, \delta=0\}}; 0 \leq u < s\}$.

On voit alors apparaître la notion de martingales.

Définition 1.2.4. Soient $X = \{X(t); t \geq 0\}$, un processus stochastique continu à droite, limité à gauche (càdlàg) et $\mathcal{F} = \{\mathcal{F}_t; t \geq 0\}$ une filtration, tous deux définis sur le même espace probabilisé. X est une martingale par rapport à la filtration \mathcal{F} si l'on a :

1. X est \mathcal{F} -adaptée ;
2. $\mathbb{E}|X(t)|$ est finie, pour $t \geq 0$;
3. $\mathbb{E}(X(t+s)|\mathcal{F}_t) = X(t)$, presque sûrement, pour t et s positifs.

Définition 1.2.5. Soit X , un processus vérifiant les propriétés 1. et 2. de la définition précédente et tel que $\mathbb{E}(X(t+s)|\mathcal{F}_t) \geq X(t)$, presque sûrement, pour t et s positifs. On dit alors que X est une sous-martingale.

Cela nous permet de voir que N est une sous-martingale et d'assurer la propriété de martingale du processus M défini comme différence $N(\cdot) - \int_0^\cdot \mathbb{1}_{\{x \geq u\}} \lambda(u) du$ par le biais de la proposition suivante.

Proposition 1.2.1. [FH90]

Soit $\{N(t); t \geq 0\}$, un processus de comptage adapté à la filtration continue à droite $\{\mathcal{F}_t; t \geq 0\}$ (i.e. telle que pour tout t , on a $\mathcal{F}_{t+} = \mathcal{F}_t$). On suppose en outre que $EN(t)$ est finie presque sûrement, pour $t \geq 0$. Alors il existe un unique processus A \mathcal{F}_t -prévisible, croissant et continu à droite, tel que

- $A(0) = 0$ presque sûrement,
- $EA(t)$ soit finie presque sûrement pour tout t ,
- $\{M(t) = N(t) - A(t); t \geq 0\}$ soit une \mathcal{F}_t -martingale continue à droite.

Définition 1.2.6. Un tel processus est appelé compensateur de la sous-martingale N . On appellera processus intensité de N , la fonction α , lorsque A s'écrit comme l'intégrale de α .

Il est possible de voir la condition (1.2) vérifiée sous certaines hypothèses, comme le montre le théorème suivant.

Théorème 1.2.1. Soit T une durée de vie absolument continue et C une variable de censure de distribution arbitraire. Soit $X = C \wedge T$, $\delta = \mathbb{1}_{\{T \leq C\}}$, et λ la fonction de risque associée à T . Définissons pour $t \geq 0$:

$$\begin{aligned} N(t) &= \mathbb{1}_{\{X \leq t; \delta=1\}}, \\ N^C(t) &= \mathbb{1}_{\{X \leq t; \delta=0\}}, \\ \mathcal{F}_t &= \sigma\{N(u), N^C(u); 0 \leq u \leq t\}. \end{aligned}$$

Alors le processus donné pour $t \geq 0$, par

$$M(t) = N(t) - \int_0^t \mathbb{1}_{\{X \geq s\}} \lambda(s) ds, \quad t \geq 0,$$

est une \mathcal{F}_t -martingale si et seulement si

$$\lambda(t) = \frac{-\frac{\partial}{\partial u} \mathbb{P}(T \geq u, C \geq t)|_{u=t}}{\mathbb{P}(T \geq t, C \geq t)}, \quad (1.3)$$

si t est tel que $\mathbb{P}(X > t) > 0$.

Démonstration : voir [FH90, p.26].

Remarque 1.2.1. On peut alors constater que l'indépendance de C et T suffit à vérifier (1.3), mais elle n'est cependant pas nécessaire.

Remarque 1.2.2. On déduit en outre que le compensateur de N est donné par $\Lambda(\cdot) = \int_0^\cdot \mathbb{1}_{\{X \geq s\}} \lambda(s) ds$ et son processus intensité s'écrit $\mathbb{1}_{\{X \geq \cdot\}} \lambda$, comme annoncé précédemment. Remarquons également, que sous les hypothèses du théorème 1.2.1, (1.2) est automatiquement vérifiée.

1.2.2 Censure à gauche

Un phénomène symétrique au précédent peut se produire : l'individu a déjà subi l'événement \mathcal{E} lors du bilan. Cette fois, on observe donc les variables aléatoires $T \vee C$ et δ , en conservant des notations similaires à celles précédemment utilisées. On parle ici de *censure à gauche*. La modélisation est similaire à la précédente.

1.2.3 Censure par intervalles

Il se peut enfin que l'information apportée par l'expérience se traduise par une appartenance de la durée de vie à un intervalle de temps, cela se produit dans le cas de suivis périodiques de patients. On parle ici de censure par intervalles (voir par exemple [Gai00]).

De façon générale, la censure se traduit par des observations effectuées sur un sous-ensemble de la tribu initialement considérée, ces sous-ensembles se distinguent suivant le type de censure abordé. Elle apporte en outre une modification dans le calcul des estimateurs, en particulier dans celui de la vraisemblance. Nous ne détaillerons pas les modifications ainsi apportées par tous les types de censure (voir par exemple [DSW00, AC96] pour les censures par intervalles et troncatures), nous privilégierons ici une explication de la provenance de l'estimateur du paramètre de régression dans le modèle de Cox et plus particulièrement de la notion de vraisemblance partielle.

1.3 Le modèle à risques proportionnels de Cox

Il est souvent possible de soupçonner une influence de certains paramètres, comme l'âge, sur la survenue de l'événement d'intérêt. Il peut s'avérer judicieux d'introduire au sein du modèle ces facteurs supposés influents, appelés par la suite variables explicatives. Le modèle de Cox introduit ces facteurs au sein de la fonction de risque.

1.3.1 Description du modèle et introduction

Notons Z la variable explicative, qui peut éventuellement être multidimensionnelle (ce qui correspondrait à plusieurs variables explicatives par individus) et dépendante du temps. Les observations sont alors constituées de $X_i, \delta_i, \{Z_i(t); 0 \leq t \leq X_i\}$, pour $1 \leq i \leq n$. Cox [Cox72] introduit la variable explicative dans la définition de la fonction de risque de la manière suivante.

Définition 1.3.1. *On dit que T satisfait un modèle à risques proportionnels de Cox lorsque l'on suppose que la fonction de risque à l'instant $t \geq 0$, conditionnellement à la variable explicative Z , est donnée par :*

$$\lambda(t; Z(\cdot)) = \lambda_0(t) \exp(Z(t)^T \beta_0), \quad (1.4)$$

où λ_0 est une fonction de risque inconnue (correspondant au cas où $\beta_0 = 0$), β_0 est un paramètre inconnu (de même dimension p que le vecteur Z) de \mathbb{R}^p .

Exemple. Un cas très courant pour la modélisation par le modèle de Cox consiste à utiliser la variable explicative pour distinguer deux catégories d'individus. Dans ce cas, Z ne peut prendre que les valeurs 0 et 1. Ce peut être le cas si l'on souhaite étudier l'effet d'un médicament pour différencier les individus traités avec un placebo des autres (pour lesquels Z vaut 1). Notons que si l'estimation conduit à une valeur de β_0 proche de 0, on pourra douter de l'influence du traitement sur la durée de vie. C'est ainsi que l'on trouve nombre de tests portant sur le modèle de Cox, visant à juger de la nullité du paramètre de régression [KM97]. Cela revient en outre à tester l'homogénéité de ces deux populations.

Remarque 1.3.1. *On constate que le modèle de Cox est décrit par deux paramètres β_0 et λ_0 , le second est fonctionnel, c'est pourquoi l'on parle de modèle semi-paramétrique.*

Remarque 1.3.2. *Comme nous le verrons plus en détails, l'estimation de β_0 par le biais de la vraisemblance partielle ne nécessite pas d'estimer λ_0 , qui sera ainsi considéré comme un paramètre de nuisance dans le modèle.*

Remarque 1.3.3. *Notons ici que la fonction de lien est exponentielle, il est possible d'introduire d'autres fonctions de lien et d'aboutir à des variations de ce modèle, l'essentiel étant de s'assurer que la fonction λ obtenue est bien une fonction de risque. Il est aussi possible d'estimer non paramétriquement cette fonction de lien, comme par exemple dans [LY95a, MS02].*

Remarque 1.3.4. *Considérons deux observations issues toutes deux du même modèle de Cox avec des variables explicatives z_1 et z_2 ne dépendant pas du temps. Si l'on quotiente les fonctions de risque, on obtient, pour $t \geq 0$,*

$$\frac{\lambda(t; z_1)}{\lambda(t; z_2)} = \exp(\beta_0(z_1 - z_2)),$$

ce qui est une quantité constante. C'est cette propriété élémentaire qui justifie l'appellation de modèle à risques proportionnels.

À partir de ce modèle, il nous faut adapter la définition de la vraisemblance à la censure des données, mais aussi à la particularité qui consiste à estimer le paramètre de régression sans estimer le paramètre de nuisance, ce qui nous conduit à définir la notion de vraisemblance partielle.

1.3.2 Estimateurs

Nous ne donnerons ici qu'une heuristique justifiant l'expression de la vraisemblance partielle, il est possible de se rapporter à [ABGK93, p.95] pour y trouver une

justification rigoureuse et précise utilisant notamment les formules de Jacod pour les processus de comptage.

La justification initiale de cette méthode fut fournie par Cox [Cox75], peu après l'article initiateur [Cox72] dans lequel figurait déjà les estimations et la notion de vraisemblance partielle. On pourra également trouver des précisions sur la vraisemblance partielle dans [FH90].

Un exemple d'heuristique pour le calcul de la vraisemblance de Cox

Reprenons ici les arguments avancés par Cox [Cox72, Cox75] et repris dans [KM97]. Supposons que les événements se produisent en des instants distincts $t_{(1)} < \dots < t_{(m)}$ (que l'on a pris soin d'ordonner). Notons $\mathcal{R}_{(t_{(i)})}$, l'ensemble des individus à risque en $t_{(i)}$, qui n'est autre que l'ensemble des individus j pour lesquels l'événement d'intérêt n'a pas encore été observé à l'instant $t_{(i)}$. On souhaite calculer la probabilité qu'un individu meure au temps $t_{(i)}$, conditionnellement aux variables explicatives $z_{(i)}$. Cela revient à considérer l'événement « un individu meurt en $t_{(i)}$, sachant qu'il y a observation de \mathcal{E} en $t_{(i)}$ » :

$$\begin{aligned} & \mathbb{P}(\text{un individu décède en } t_{(i)} \mid \text{il y a une mort en } t_{(i)}) \\ = & \frac{\mathbb{P}(\text{un individu décède en } t_{(i)} \mid \text{il a survécu jusqu'à } t_{(i)})}{\mathbb{P}(\text{il y a une mort en } t_{(i)} \mid \text{il a survécu jusqu'à } t_{(i)})} \\ \approx & \frac{\lambda(t_{(i)})}{\sum_{l \in \mathcal{R}(t_{(i)})} \lambda(t_{(l)})} \\ = & \frac{\exp(z_{(i)}\beta)}{\sum_{l \in \mathcal{R}(t_{(i)})} \exp(z_{(l)}\beta)}, \end{aligned}$$

qui est la contribution de l'individu (i) à la vraisemblance de Cox (qualifiée dans [Cox72] de conditionnelle, puis intégrée dans un cadre plus général comme étant une vraisemblance partielle). Le principal intérêt de cette formule est le fait qu'elle ne fasse pas intervenir le paramètre de nuisance. On obtient alors une première expression de la log-vraisemblance C_n :

$$C_n(\beta) = \sum_{i=1}^n \left\{ z_{(i)}\beta - \log \left(\sum_{l \in \mathcal{R}(t_{(i)})} \exp(z_{(l)}\beta) \right) \right\}.$$

On prendra alors $\hat{\beta}_n = \operatorname{argmax} C_n(\beta)$ pour estimateur de β_0 .

La vraisemblance partielle de Cox vue comme une vraisemblance profilée

Remarquons qu'il est possible de faire apparaître cette vraisemblance de façon naturelle par la biais des vraisemblances *profilées* (voir par exemple [KM97, MvdV00, vdV02]).

Supposons les variables explicatives indépendantes du temps et calculons la vraisemblance du couple de paramètres de $(x_i, \delta_i, z_i)_{1 \leq i \leq n}$:

$$l_n(\beta_0, \lambda_0) = \prod_{i=1}^n \lambda_0(t_i)^{\delta_i} (\exp(\beta_0 z_i))^{\delta_i} \exp(-\Lambda_0(t_i) \exp(\beta_0 z_i)).$$

Dans un premier temps, on considère cette vraisemblance comme fonction de λ_0 uniquement. Pour maximiser cette dernière, on suppose que :

$$\Lambda_0(t) = \sum_{i=1}^n \mathbb{1}_{\{t_i \leq t\}} \lambda_i, \quad \lambda_i \in \mathbb{R}_+.$$

Le problème consiste alors à déterminer les $\tilde{\lambda}_i, 1 \leq i \leq n$, qui maximisent la vraisemblance, qui s'exprime désormais comme suit

$$l_n(\beta, \lambda_1, \dots, \lambda_n) = \prod_{i=1}^n \lambda_i^{\delta_i} (\exp(\beta z_i))^{\delta_i} \exp\left(-\sum_{\{j; t_j \leq t_i\}} \lambda_j \exp(\beta z_i)\right).$$

Alors la dérivée de la log-vraisemblance par rapport à λ_{i_0} s'écrit :

$$\frac{d \log l_n}{d \lambda_{i_0}}(\beta, \lambda_1, \dots, \lambda_n) = \sum_{i=1}^n \frac{\delta_{i_0}}{\lambda_{i_0}} - \sum_{i=1}^n \sum_{\{j; t_j \leq t_i\}} \mathbb{1}_{\{j=i_0\}} \exp(\beta z_i),$$

ce qui donne, en inversant les sommes sur i et j dans la seconde partie de l'expression, et en annulant cette dérivée partielle :

$$\begin{aligned} \frac{\delta_{i_0}}{\tilde{\lambda}_{i_0}} &= \sum_{i=1}^n \mathbb{1}_{\{t_{i_0} \leq t_i\}} \exp(\beta z_i) \\ &= \sum_{\{i; t_i \geq t_{i_0}\}} \exp(\beta z_i). \end{aligned}$$

On définit alors la vraisemblance profilée par :

$$\begin{aligned} l_n(\beta) &= l_n(\beta, \tilde{\lambda}_1, \dots, \tilde{\lambda}_n) \\ &= \prod_{i=1}^n \left(\frac{\exp(\beta z_i)}{\sum_{\{j; t_j \geq t_i\}} \exp(\beta z_j)} \right)^{\delta_i} \\ &\times \exp\left(-\sum_{j=1}^n \frac{\sum_{i=1}^n \mathbb{1}_{\{t_j \leq t_i\}} \exp(\beta z_i)}{\sum_{\{l; t_l \geq t_j\}} \exp(\beta z_l)}\right) \\ &= \prod_{i=1}^n e^{-1} \times \left(\frac{\exp(\beta z_i)}{\sum_{\{j; t_j \geq t_i\}} \exp(\beta z_j)} \right)^{\delta_i}. \end{aligned}$$

On retrouve alors l'expression habituelle, à un facteur multiplicatif près. C'est elle que l'on maximisera en fonction de β .

Propriétés

Il est possible de trouver une expression de cette vraisemblance profilée pour divers phénomènes de censures dans [vdV02], ainsi que des propriétés de convergence s'y rattachant. On pourra également se référer à [Won86] pour les propriétés des estimateurs de maximum de vraisemblance partielle, et enfin à [ABGK93] pour une vision de ces estimations par le biais des processus de comptage. L'une des premières approches concernant les propriétés des estimateurs fut apportée par [Tsi81], puis peu après par [AG82] à l'aide de techniques utilisant les martingales.

Précisons dans un premier temps les notations utilisées tout au long de ce mémoire (semblables à celles utilisées par [AG82]) ainsi que le cadre de notre étude. Considérons un échantillon X_1, \dots, X_n de variables aléatoires telles que X_i soit la plus petite des valeurs prises par T_i et C_i , avec C_i la variable de censure, T_i la durée de vie. Les données sont supposées censurées à droite et les observations sont $X_i, \delta_i, \{Z_i(s); 0 \leq s \leq X_i\}, i = 1, \dots, n$, où les Z_i sont les variables explicatives et δ_i , les fonctions indicatrices d'observation de \mathcal{E} . Supposons en outre que la fonction de risque relative à T_i est donnée par (1.4). Considérons Y_i le processus défini par

$$Y_i(t) = \mathbb{1}_{\{X_i \geq t\}},$$

$Y_i(t)$ indique si l'individu i est à risque à l'instant t , c'est-à-dire s'il n'a pas subi \mathcal{E} . $(Y_i)_{i \geq 1}$ est alors un processus prévisible. On note N_i , le processus défini par

$$N_i(t) = \mathbb{1}_{\{\delta_i=1, X_i \leq t\}}.$$

Notons que N_i change de valeur uniquement lorsque Y_i vaut 1. Afin d'obtenir une estimation de β , on maximise la log-vraisemblance partielle (obtenue dans [Cox75]), et donnée par

$$C_n(\beta, \tau) = \sum_{i=1}^n \int_0^\tau \beta^T Z_i(s) dN_i(s) - \int_0^\tau \log \left(\sum_{i=1}^n Y_i(s) e^{\beta^T Z_i(s)} \right) d\bar{N}(s),$$

avec $\bar{N} = \sum_{i=1}^n N_i$ et τ , la durée d'observation, satisfaisant $\tau \leq +\infty$. Remarquons ici que $\hat{\beta}_n(\tau) = \operatorname{argmax}_{\beta \in \mathbf{R}^p} C_n(\beta, \tau)$ conduit à un estimateur qui dépend du temps τ . Ce dernier coïncide avec l'estimateur de Cox lorsque $\tau = +\infty$.

Afin de simplifier la manipulation matricielle, nous utiliserons la notation suivante pour les produits de matrice (qui se résume à une simple puissance dans le cas unidimensionnel) :

$$\begin{aligned} U^{\otimes 0} &= 1, \\ U^{\otimes 1} &= U, \\ U^{\otimes 2} &= U^T U. \end{aligned}$$

Pour simplifier les formulations des équations relatives à la vraisemblance de

Cox, nous définissons les processus suivants, pour β dans \mathbb{R}^p et t dans $[0, \tau]$,

$$\begin{aligned} S_n^{(m)}(\beta, t) &= \frac{1}{n} \sum_{l=1}^n Z_l(t)^{\otimes m} Y_l(t) e^{\beta^T Z_l(t)}, \quad m \in \{0, 1, 2\}, \\ E_n(\beta, t) &= \frac{S_n^{(1)}(\beta, t)}{S_n^{(0)}(\beta, t)}, \\ V_n(\beta, t) &= \frac{S_n^{(2)}(\beta, t)}{S_n^{(0)}(\beta, t)} - E_n(\beta, t)^{\otimes 2}. \end{aligned}$$

En dérivant C_n par rapport à β , on obtient les formulations suivantes pour la fonction score U_n et la matrice d'information de Fisher \mathcal{I}_n :

$$\begin{aligned} U_n(\beta, t) &= \frac{1}{n} \frac{\partial C_n}{\partial \beta}(\beta, t) \\ &= \frac{1}{n} \sum_{i=1}^n \int_0^t Z_i(s) dN_i(s) - \frac{1}{n} \int_0^t \frac{S_n^{(1)}(\beta, s)}{S_n^{(0)}(\beta, s)} d\bar{N}(s), \\ \mathcal{I}_n(\beta, t) &= -\frac{\partial U_n}{\partial \beta}(\beta, t) \\ &= -\frac{1}{n} \frac{\partial^2 C_n}{\partial \beta \partial \beta^T}(\beta, t) \\ &= \frac{1}{n} \int_0^t V_n(\beta, s) d\bar{N}(s). \end{aligned}$$

Remarque 1.3.5. La variable β sera parfois omise dans les notations lorsque $\beta = \beta_0$, ceci afin de simplifier certaines expressions.

Les hypothèses suivantes permettent d'établir la convergence et la normalité asymptotique des principaux estimateurs relatifs au modèle de Cox.

Hypothèses 1. [AG82]

H1 - $\int_0^\tau \lambda_0(t) dt$ est fini.

H2 - Il existe un voisinage \mathcal{B} de β_0 , des fonctions $s^{(0)}$, $s^{(1)}$ et $s^{(2)}$ définies sur $\mathcal{B} \times [0, \tau]$ telles que, pour $j = 0, 1, 2$,

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|S_n^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \xrightarrow{P} 0,$$

lorsque $n \rightarrow +\infty$.

H3 - En conservant les notations précédentes, on définit

$$e = \frac{s^{(1)}}{s^{(0)}} \quad \text{et} \quad v = \frac{s^{(2)}}{s^{(0)}} - e^{\otimes 2}.$$

Pour β dans \mathcal{B} , $t \in [0, \tau]$,

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s^{(0)}(\beta, t), \quad s^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta \partial \beta^T} s^{(0)}(\beta, t),$$

$s^{(0)}(\cdot, t)$, $s^{(1)}(\cdot, t)$, et $s^{(2)}(\cdot, t)$ sont des fonctions continues sur \mathcal{B} , uniformément en $t \in [0, \tau]$; $s^{(0)}$ est bornée inférieurement par une constante strictement positive sur $\mathcal{B} \times [0, \tau]$, et

$$\mathcal{I}(\beta_0) = \int_0^\tau v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt$$

est une matrice définie positive.

Sous ces hypothèses, le théorème suivant résume le comportement de l'estimateur $\hat{\beta}_n$ du paramètre de régression par la méthode du maximum de vraisemblance partielle. Dans la suite de ce mémoire, on s'y référera fréquemment. En effet, les propriétés obtenues sont nécessaires à l'approche séquentielle. Toutefois, notons que les hypothèses émises sur les données au cours du chapitre 4, bien que voisines de celles d'Andersen et Gill [AG82], présentent des différences notables. En particulier, l'hypothèse (H2) sera formulée sous une forme moins forte.

Théorème 1.3.1.

Sous les hypothèses (H1) à (H3), les propriétés suivantes sont vérifiées.

1. (Convergence) $\hat{\beta}_n \xrightarrow{P} \beta_0$, lorsque $n \rightarrow +\infty$;
2. (Normalité asymptotique) $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} \mathcal{N}(0, \mathcal{I}(\beta_0)^{-1})$, lorsque $n \rightarrow +\infty$;
3. l'estimateur de la variance asymptotique $\hat{\sigma}_n^2 = \mathcal{I}_n(\hat{\beta}_n, \tau)^{-1}$ converge en probabilité vers $\mathcal{I}(\beta_0)^{-1}$, lorsque $n \rightarrow +\infty$.

La démonstration du théorème qui précède figure dans [AG82].

Remarque 1.3.6. Une étape clé de la preuve du théorème 1.3.1 consiste à montrer que sous (H1)-(H3), on a :

$$\sqrt{n}U_n(\beta_0, \tau) \xrightarrow{D} \mathcal{N}(0, \mathcal{I}(\beta_0)), \quad \text{lorsque } n \rightarrow +\infty.$$

1.4 Procédures séquentielles d'arrêt

n considère un échantillon X_1, \dots, X_n issu d'une distribution de probabilité inconnue P sur $(\mathbb{X}, \mathcal{B})$, où \mathbb{X} est un espace euclidien muni de sa tribu borélienne

\mathcal{B} . On suppose que la donnée de β et η (paramètre de nuisance) permet de déterminer \mathcal{P} de manière unique. Lorsque β et η appartiennent à des espaces de dimension finies, le modèle est alors paramétrique. Dans le cas contraire, on parle de modèle semi-paramétrique, voire non paramétrique. On pose $\theta = (\beta, \eta)$. La loi de probabilité \mathcal{P}_θ cherchée appartient à un sous-ensemble \mathbb{P} (le modèle) de l'ensemble \mathbb{M} de toutes les mesures de probabilités sur $(\mathbb{X}, \mathcal{B})$. On suppose que l'estimation de β ne requiert pas celle de η , c'est bien le cas lorsque l'on considère le modèle de Cox sous les hypothèses d'Andersen et Gill (le paramètre η n'est autre que la fonction de risque de base λ_0). On se place dans le cadre où chaque nouvelle donnée est coûteuse (les critères de coût peuvent être humains, financiers ou de temps), et le statisticien souhaite mener une étude optimale, dans le sens où il veut concilier qualité de l'estimation et coût minimal.

Si le statisticien juge que la quantité de données dont il dispose est insuffisante pour garantir la qualité de l'estimation, il doit, pour affiner son estimation, introduire de nouvelles données. Ces nouvelles données peuvent avoir un coût conséquent car elles peuvent être le fruit d'une nouvelle enquête, de nouvelles expérimentations, etc. D'autre part, définir la qualité de l'estimation nécessite de définir un critère quantitatif. Finalement, c'est en prenant en compte à la fois le coût du recueil des données et le critère quantitatif de qualité de l'estimation, que le statisticien définira une règle d'arrêt réalisant un compromis entre le coût total de l'étude et l'objectif de qualité d'estimation fixé *a priori*.

Une fois la règle d'arrêt établie, à chaque nouvel apport de donnée ou de groupe de données, on peut vérifier si la règle d'arrêt est satisfaite. L'étude cesse lorsque la règle d'arrêt est satisfaite. L'ensemble de la procédure conduisant à une estimation est appelée procédure séquentielle d'arrêt, l'estimateur résultant est appelé estimateur séquentiel.

Fait important, un estimateur sera en général indexé par une taille d'échantillon aléatoire. De fait, il n'est pas évident qu'il hérite des mêmes propriétés que celles de l'estimateur non séquentiel (par exemple la convergence et la normalité asymptotique). Il s'avère alors nécessaire de procéder à une étude du prolongement de ces propriétés au cas de tailles d'échantillons fournies pour des règles séquentielles d'arrêt. Nous allons voir dans ce qui suit diverses procédures séquentielles d'arrêt ainsi que les principales propriétés qui les caractérisent.

1.4.1 Préoccupations menant à diverses procédures d'arrêt

Pour un panel d'observations donné, il serait souhaitable d'associer à l'estimation obtenue une quantité représentative de la qualité de l'observation. On peut ainsi stopper l'étude (et limiter son coût) dès qu'un seuil (défini au préalable) est atteint. Cette qualité peut être définie naïvement par la distance entre le paramètre et son estimateur, *i.e.* par une fonction dite d'erreur. Puisque liée au paramètre inconnu, l'erreur n'est cependant pas directement observable. Il est possible d'étudier la fonction de risque qui lui est attachée (elle est définie par l'espérance de l'erreur)

et d'explorer des techniques bayésiennes visant à la minimiser. Une alternative vise à étudier la probabilité que le paramètre soit dans l'intervalle de confiance obtenu par les propriétés de normalité asymptotique de l'estimateur. L'une ou l'autre de ces techniques amène à la définition d'un critère de suffisance pour stopper l'étude : c'est la règle d'arrêt. Elle définit alors un temps d'arrêt pour lequel il se révèle crucial d'obtenir des résultats asymptotiques (nous précisons plus tard le sens que l'on peut donner à ce mot). Dans cette partie introductive, nous verrons quelques préoccupations menant à définir des règles d'arrêt, ainsi que les propriétés qu'il est intéressant d'étudier à leur sujet. Notons que le lecteur intéressé pourra par exemple se référer à [Dur97a, Dur97b] où il est question d'estimations séquentielles pour un modèle de régression linéaire.

On peut distinguer trois problématiques générales qui correspondent à trois préoccupations du statisticien [Gov86]. Pour un n -échantillon de variables aléatoires de densité $f(\cdot; \beta)$, on souhaite estimer le paramètre β . À cette fin, on se donne un estimateur $\hat{\beta}_n$ de β ainsi qu'une fonction de perte $r(\hat{\beta}_n; \beta)$. Le coût de l'étude est ensuite donné par $C(n)$, où n est le nombre d'observations à ajouter pour vérifier le critère souhaité. $C(n)$ peut en pratique être exprimé sous la forme $c_0 + cn$, où c_0 est le coût initial de l'étude, et c représente le coût d'une nouvelle observation [GMS97]. Il s'agit alors d'un coût déterministe. En sus des choix de la règle d'arrêt et d'un estimateur pertinent du paramètre β , les stratégies du statisticien sont multiples.

1. Il est effectivement possible de minimiser le risque $\mathbb{E}(r(\hat{\beta}_n; \beta))$ en le soumettant à une contrainte de la forme suivante $\mathbb{E}C(n) \leq n_{\max}$. Cela correspond à se donner un coût moyen maximal pour l'expérience et optimiser ensuite la procédure (limitée par ce coût).
2. La stratégie inverse est également possible : on peut minimiser le coût moyen de l'expérience $\mathbb{E}C(n)$ tout en s'assurant que la fonction de perte ne dépasse pas une borne fixée.
3. Enfin, une dernière technique permet d'aborder le problème de façon bilatérale grâce à l'introduction de la fonction de coût dans la fonction de perte, qui pourra être de la forme $\mathbb{E}(r(\hat{\beta}_n; \beta) + C(n))$. C'est par exemple la méthode utilisée dans [GD02], dans le cas de l'estimation de la variance.

Ces préoccupations aboutissent à diverses méthodes séquentielles. Notons parmi elles, les procédures basées sur l'amplitude des intervalles de confiance. Cette technique consiste à rechercher la taille minimale de l'échantillon sous contrainte de l'appartenance du paramètre à un intervalle de confiance de longueur fixée, pour un niveau de confiance donné. L'un des articles initiateurs de ces techniques fut celui de [CR65]. Quelques années plus tard, [SW76] étudia les différentes propriétés des estimateurs séquentiels. Il distingue trois types d'études asymptotiques basées sur la probabilité que l'estimateur ne soit pas dans l'intervalle de confiance séquentiel associé, ainsi que la longueur de cet intervalle. Le comportement asymptotique le plus couramment considéré est alors celui obtenu en faisant tendre la taille de l'intervalle de confiance vers 0.

Les choix d'un critère d'arrêt et d'un estimateur associé permettent de déterminer entièrement une procédure séquentielle. Cependant il peut s'avérer, qu'avec un tel

choix effectué, aucune procédure ne réponde aux demandes de l'utilisateur. Nous illustrons ces propos à l'aide d'un exemple simple tiré de [GMS97], que l'on peut aussi retrouver dans [Gov86]. Nous définirons par la suite deux règles d'arrêt.

Exemple [GMS97, GD02, Gov86]

Plaçons-nous dans le cas de variables aléatoires indépendantes et identiquement distribuées X_1, \dots, X_n suivant une loi gaussienne de moyenne μ et de variance σ^2 . Considérons la fonction de perte L donnée par l'erreur quadratique, et la fonction de coût C associée précédemment définie par $C(n) = c_0 + cn$.

Nous souhaitons dans un premier temps estimer la moyenne μ . L'estimateur le plus naturel qui lui est associé est donné par la moyenne empirique \bar{X}_n des variables $X_i, 1 \leq i \leq n$. Chaque étape de la procédure séquentielle vise à minimiser le risque donné par

$$C(n) + \mathbb{E}L(\bar{X}_n - \mu) = c_0 + cn + \frac{\sigma^2}{n},$$

ce qui donne, après un calcul simple,

$$n^* = \inf \left\{ n \geq 1; n^2 \geq \frac{\sigma^2}{c} \right\}.$$

n^* dépend de la variance inconnue σ^2 , qu'il est néanmoins possible de remplacer par son estimateur sans biais (la variance empirique corrigée) afin d'obtenir une règle d'arrêt observable notée N . En pratique, elle est définie par :

$$N = \inf \left\{ n \geq 1; n^2 \geq \frac{S_n^2}{c} \right\},$$

où $S_n^2 = (n-1)^{-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$. Il est alors possible de considérer \bar{X}_N comme un estimateur ponctuel de la moyenne μ et de montrer ses propriétés [GMS97]. Il est également possible de considérer (σ, μ) comme paramètre et de mener une étude similaire. Reprenons les notations de l'exemple précédent. Supposons dès lors que l'on souhaite établir un intervalle de confiance séquentiel pour μ . Un intervalle naturel basé sur la statistique $T_n = \sqrt{n}(\bar{X}_n - \mu)/\sigma$ nous est donné par

$$[\bar{X}_n - n^{-1/2}\sigma u_\alpha, \bar{X}_n + n^{-1/2}\sigma u_\alpha], \quad (1.5)$$

α est un niveau de confiance fixé, u_α est le quantile d'ordre $1 - \alpha/2$ d'une loi normale centrée réduite.

Si σ^2 est connue, il suffit de poser

$$n^* = \inf \left\{ n \geq 1; n \geq d^{-2}\sigma^2 u_\alpha^2 \right\}, \quad (1.6)$$

pour s'assurer que l'amplitude de l'intervalle choisi est inférieure à $2d$. On aura en outre la propriété

$$\mathbb{P}(\bar{X}_{n^*} - d \leq \mu \leq \bar{X}_{n^*} + d) \geq 1 - \alpha. \quad (1.7)$$

Étudions désormais le cas où la variance est inconnue. Pour n fixé, l'amplitude de l'intervalle de confiance donnée par (1.5) peut ou non être inférieure à $2d$, selon la

position de n par rapport à n^* . Puisque la vraie valeur de n^* dépend de σ^2 , elle est inconnue, et il est donc impossible de créer une procédure dont on puisse affirmer la qualité (et donc (1.7)), pour toute valeur de σ^2 , avec une taille d'échantillon fixée.

De la même façon, si l'on considère l'intervalle de confiance donné par

$$[\bar{X}_n - d, \bar{X}_n + d],$$

la probabilité pour que μ soit dans cet intervalle est donnée par

$$2\Phi\left(\frac{\sqrt{nd}}{\sigma}\right) - 1, \quad (1.8)$$

où Φ est la fonction de répartition d'une loi normale centrée réduite. Il est alors impossible de connaître la position de (1.8) par rapport à $1 - \alpha$, puisque cela dépend de celle de $n^{1/2}d\sigma^{-1}$ par rapport à u_α . Il nous est donc impossible de fournir, pour d fixé, une règle d'arrêt qui vérifie les conditions souhaitées, pour tout $\sigma > 0$.

1.4.2 Définitions générales

Un article précurseur pour la théorie asymptotique des estimateurs séquentiels est sans doute [Ans50], dans le cas paramétrique. Quelques années plus tard, Anscombe [Ans52, Ans53] établit la notion de continuité uniforme en probabilité et fournit des conditions suffisantes pour établir la normalité asymptotique d'une suite d'estimateurs indexée par une suite de variables aléatoire entières et positives (le lecteur intéressé pourra se référer à l'Annexe C qui concerne la condition d'Anscombe et quelques applications en analyse séquentielle, ainsi qu'à [GMS97] qui présente de nombreuses méthodes séquentielles et les techniques mathématiques permettant d'établir leurs propriétés).

Le cadre de notre étude se caractérise principalement par le fait que l'acquisition d'une nouvelle donnée entraîne un coût non négligeable qu'on souhaiterait minimiser. Il convient alors de modéliser le concept de « bonne estimation » (*i.e.* suffisante à l'arrêt de l'étude). Cette notion est relative aux besoins du statisticien, nous avons précédemment entrevu l'existence de plusieurs types de règles d'arrêt et bien sûr la notion d'estimation y est intrinsèquement liée. Nous nous plaçons ici dans le cas d'estimations séquentielles. Notre but consiste donc à estimer β_0 , paramètre de régression, par le biais de son estimateur pour n observations, noté $\hat{\beta}_n$. On suppose que l'utilisateur a fixé une conduite d'arrêt, matérialisée mathématiquement par l'existence d'une variable aléatoire N_d , qui donne, pour un niveau de précision $d > 0$ fixé par le statisticien, la taille d'échantillon minimale pour laquelle le niveau de précision d est atteint. L'estimateur obtenu est alors $\hat{\beta}_{N_d}$. Si l'estimateur non séquentiel $\hat{\beta}_n$ possède des propriétés de convergence, il se révèle pertinent d'étudier les mêmes propriétés pour l'estimateur séquentiel $\hat{\beta}_{N_d}$, qui est alors un estimateur indexé par une taille aléatoire d'échantillon. Ceci est d'autant plus important que les critères d'arrêt sont souvent basés sur des propriétés des estimateurs non séquentiels. On suppose maintenant que la précision d'estimation croît lorsque d tend vers 0.

La notion de régularité permet de constater une certaine analogie entre les comportements des suites $(N_d)_{d>0}$ et $(n_d)_{d>0}$, où la seconde suite est une suite de variables d'arrêt déterministe, en général inaccessible car dépendante de paramètres inconnus du modèle. Appeler $(N_d)_{d>0}$ ou $(n_d)_{d>0}$ « suites » est ici un abus de langage qui est levé si l'on indexe ces familles par des suites $(d_k)_{k\geq 1}$ décroissant vers 0. L'asymptotique résulte alors de « k tend vers l'infini », ou de « d tend vers 0 ».

Définition 1.4.1. *La suite $(N_d)_{d>0}$ est régulière s'il existe une suite déterministe d'entiers $(n_d)_{d>0}$ tendant vers l'infini, lorsque $d \rightarrow 0$ et vérifiant*

$$\frac{N_d}{n_d} \xrightarrow{P} 1, \text{ lorsque } d \rightarrow 0.$$

De plus, $(N_d)_{d>0}$ est dite asymptotiquement efficace si :

$$\frac{IEN_d}{n_d} \rightarrow 1, \text{ lorsque } d \rightarrow 0.$$

On suppose que l'on dispose en outre d'un intervalle de confiance I_{N_d} pour une taille d'échantillon N_d . Pour un niveau de confiance α donné, il est souhaitable que l'intervalle de confiance séquentiel vérifie des propriétés similaires à celles habituellement supposées dans le cadre général d'intervalles de confiance, comme par exemple la normalité asymptotique. Elle se traduit dans le cas d'intervalles de confiance séquentiels par la notion de convergence asymptotique.

Définition 1.4.2. *On dit que I_{N_d} , intervalle de confiance séquentiel associé à la taille d'échantillon N_d , est asymptotiquement convergent au niveau de confiance α , si, lorsque $d \rightarrow 0$,*

$$IP(\beta_0 \in I_{N_d}) \rightarrow 1 - \alpha.$$

Remarque 1.4.1. *La normalité asymptotique de l'estimateur $\hat{\beta}_{N_d}$ entraîne la propriété précédente.*

Il est également possible de juger du biais introduit par la règle séquentielle d'arrêt par l'étude de la quantité $IEN_d - n_d$ (loi asymptotique, vitesse de convergence, bornitude, ...), appelée coût de l'ignorance [DG00].

Dans le cas particulier du modèle semi-paramétrique, nous verrons que les résultats obtenus dans le cadre séquentiel ne présupposent pas de choix particulier d'une règle d'arrêt, elle doit cependant vérifier certaines propriétés. C'est ainsi que nous sommes amenés à définir un ensemble \mathcal{A} de règles d'arrêt pour lesquelles nous disposons d'un théorème (voir chapitre 3) permettant d'étendre les propriétés non séquentielles des estimateurs au cadre séquentiel.

Définition 1.4.3. *On note \mathcal{A} , l'ensemble de règles séquentielles d'arrêt $(N_d)_{d>0}$ qui vérifient*

- N_d existe presque sûrement, pour tout d réel strictement positif;
- $(N_d)_{d>0}$ est une suite régulière de règles d'arrêt.

Nous allons voir dans un premier temps qu'il n'est pas nécessaire pour une règle d'arrêt d'être régulière pour vérifier la propriété de convergence asymptotique, ce qui est illustré par la méthode en deux étapes de Stein [Ste45].

Exemple (suite)

Une solution au problème abordé précédemment fût apportée par Stein [Ste45], qui considéra une procédure en deux étapes pour obtenir des règles séquentielles d'arrêt satisfaisantes. Cette règle d'arrêt présente un intérêt particulier pour nous, puisque nous verrons que bien qu'étant asymptotiquement convergente, elle n'est pas régulière. Elle apporte donc une réponse à la question de la nécessité de la régularité des règles d'arrêt. Nous verrons par la suite l'importance de la régularité dans le cadre des résultats obtenus au chapitre 3.

Le problème consiste à trouver une taille d'échantillon optimale telle que l'on ait simultanément $\mathbb{P}(\mu \in I_n) \geq 1 - \alpha$ et $|I_n| = 2d$, où I_n est l'intervalle de confiance séquentiel considéré. On sait que

$$\mathbb{P}(\mu \in I_n) = 2\Phi\left(\frac{\sqrt{n}d}{\sigma}\right) - 1,$$

la taille optimale n_d de l'échantillon est donnée par le plus petit entier supérieur à la quantité $u_\alpha^2 \sigma^2 d^{-2}$, où $\Phi(u_\alpha) = 1 - \alpha/2$ (n_d est inconnue). On sait que la variable donnée par $S_n^{-1}(\bar{X}_n - \mu)$ suit une loi de Student à $(n - 1)$ degrés de liberté, notée τ_{n-1} . Décrivons alors la procédure de Stein : pour m entier supérieur à 2, on note a_{m-1} , le quantile d'ordre $\alpha/2$ d'une loi τ_{m-1} , et si :

1. $a_{m-1}^2 S_m^2 d^{-2} + 1 < m$, on décide de considérer un m -échantillon ;
2. sinon, on prend les $\lceil a_{m-1}^2 S_m^2 d^{-2} \rceil + 1 - m$ observations suivantes, où $\lceil \cdot \rceil$ désigne la partie entière.

On obtient alors un échantillon de taille N_d donnée par

$$N_d = \max \left\{ m; \left\lceil \frac{a_{m-1}^2 S_m^2}{d^2} \right\rceil + 1 \right\},$$

l'intervalle de confiance associé est alors donné par $I_{N_d} = [\bar{X}_{N_d} - d, \bar{X}_{N_d} + d]$. Il est ensuite possible de montrer [GMS97, p.154] que N_d vérifie, pour tous μ et σ^2 ,

$$\begin{aligned} \mathbb{P}(\mu \in I_{N_d}) &\geq 1 - \alpha, \\ \lim_{d \rightarrow 0} \mathbb{P}(\mu \in I_{N_d}) &= 1 - \alpha, \end{aligned}$$

et bien sûr, la taille de l'intervalle de confiance est de $2d$. Par ailleurs, on peut voir que la règle d'arrêt ainsi définie n'est pas régulière puisqu'elle induit un choix de l'entier m définissant la taille de l'échantillon

$$\lim_{d \rightarrow 0} \mathbb{E} \left(\frac{N_d}{n_d} \right) = \frac{a_{m-1}^2}{a^2}.$$

Nous avons vu qu'il est possible de construire une procédure d'arrêt non régulière qui soit convergente, ce qui induit en particulier que la régularité d'un estimateur

séquentiel n'est pas nécessaire. Toutefois, quelques modifications de ces règles d'arrêt conduisant à des règles d'arrêt régulières, on peut également s'interroger sur d'autres propriétés [GMS97, MD97]. Dans la suite de ce chapitre, nous présentons deux règles d'arrêt pour lesquelles nous établirons les principales propriétés au chapitre suivant et nous verrons ainsi qu'il est possible de les utiliser dans le cadre du modèle de Cox.

1.4.3 Critères d'arrêt dans le cadre semi-paramétrique

Rappelons que nous considérons ici des variables aléatoires X_1, \dots, X_n , et que la données des paramètres β_0 et η permettent de définir entièrement leur distribution. On suppose qu'il est possible d'estimer β_0 sans disposer d'estimateur de η . Notre but, au cours de cette partie, est de décrire deux règles possibles d'arrêt pour l'estimation séquentielle de β_0 . Dans un premier temps, nous décrivons les propriétés que doit satisfaire l'estimateur non séquentiel $\hat{\beta}_n$ (n représente la taille de l'échantillon) de β_0 .

Hypothèses sur le modèle

Il apparaît clairement nécessaire de supposer quelques résultats élémentaires de convergence de l'estimateur considéré, par mesure de cohérence dans la définition des règles d'arrêt, c'est pourquoi nous sommes amenés à faire ces hypothèses.

Hypothèses 2.

- Convergence de $\hat{\beta}_n$: $\hat{\beta}_n \xrightarrow{P} \beta_0$, lorsque $n \rightarrow \infty$.
- Normalité asymptotique : $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$, lorsque $n \rightarrow \infty$, avec $\sigma > 0$.
- Convergence de $\hat{\sigma}_n^2$: on dispose d'un estimateur de la variance asymptotique, noté $\hat{\sigma}_n^2$, lui aussi convergeant, i.e. tel que $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$, lorsque $n \rightarrow +\infty$.

Intervalles séquentiels de confiance

Une première approche se base sur des estimations séquentielles de l'intervalle de confiance. L'utilisateur fixe une quantité strictement positive $2d$ qui correspond à la largeur maximale de l'intervalle de confiance obtenu. Nous verrons au chapitre 2 qu'il est possible de choisir une valeur de d relative à celle de β_0 par le biais de son estimateur.

Soit α , un seuil de confiance fixé. Par les hypothèses de normalité asymptotique, on sait que :

$$\forall u \in \mathbb{R}, \mathbb{P}_\theta \left(\left| \frac{\sqrt{n}(\hat{\beta}_n - \beta_0)}{\sigma} \right| \leq u \right) \rightarrow 2\Phi(u) - 1,$$

où Φ est la fonction de répartition d'une loi normale centrée réduite.

En particulier, si l'on choisit u_α tel que $\Phi(u_\alpha) = 1 - \alpha/2$, l'intervalle de confiance asymptotique pour β_0 au niveau de confiance $1 - \alpha$ s'écrit alors :

$$\left[\hat{\beta}_n - \frac{u_\alpha \sigma}{\sqrt{n}}; \hat{\beta}_n + \frac{u_\alpha \sigma}{\sqrt{n}} \right],$$

il est donc d'amplitude $\frac{2u_\alpha \sigma}{\sqrt{n}}$.

Si l'on souhaite connaître la taille minimale de l'échantillon nécessaire pour obtenir un intervalle de confiance asymptotique de niveau de confiance $1 - \alpha$ et d'amplitude inférieure à $2d$, il nous suffit de prendre une taille d'échantillon n telle que $n \geq \frac{u_\alpha \sigma^2}{d^2}$.

Remarque 1.4.2. *Pour que la variable d'arrêt ainsi définie soit pertinente, et étant donné qu'elle se base sur un résultat asymptotique, il est préférable d'imposer une valeur minimale pour n , c'est pourquoi on prendra :*

$$n_d = \min \left\{ n; n \geq n_0 \text{ et } n \geq \frac{(u_\alpha \sigma)^2}{d^2} \right\}.$$

En pratique, on dispose seulement d'un estimateur $\hat{\sigma}_n^2$ de la variance asymptotique; on définit donc par analogie avec le raisonnement qui précède la variable d'arrêt N_d par

$$N_d = \min \left\{ n \geq n_0; n \geq \frac{(u_\alpha \hat{\sigma}_n)^2}{d^2} \right\}.$$

En quelque sorte, N_d peut être vu comme un estimateur de n_d .

Remarque 1.4.3. *On ne peut pas contrôler le comportement de l'estimateur de la variance asymptotique pour de faibles tailles d'échantillons; en particulier, il peut être très proche de zéro. Pour pallier ce problème, on considère une suite $(\varepsilon_n)_{n \geq 1}$ de réels strictement positifs qui décroît vers zéro, et on substitue $\max(\varepsilon_n, \hat{\sigma}_n) = \hat{\sigma}_n^*$ à $\hat{\sigma}_n$ dans la règle d'arrêt.*

Remarque 1.4.4. *On a toujours le résultat : $\hat{\sigma}_n^{2*} \xrightarrow{P} \sigma^2$ (comme nous le verrons plus tard). Dans [GMS97], une méthode similaire est utilisée, remarquons toutefois qu'on choisit ici une suite $(\varepsilon_n)_{n \geq 1}$ qui varie suivant la taille de l'échantillon, et non suivant l'amplitude de l'intervalle.*

Définition 1.4.4. *Soit n_0 , un entier strictement positif fixé. Soit $d > 0$ et $\alpha \in]0, 1[$, la variable d'arrêt est définie par*

$$N_d = \min \left\{ n \geq n_0; n \geq \frac{(u_\alpha \hat{\sigma}_n^*)^2}{d^2} \right\}. \quad (1.9)$$

Remarque 1.4.5. *Avec une telle définition, $(N_d)_{d>0}$ est presque sûrement décroissante.*

Critère basé sur la minimisation d'un risque

Supposons désormais que notre règle d'arrêt soit basée sur la minimisation d'une fonction de risque constituée d'une erreur r portant sur l'estimation et d'un coût C défini par $C(n) = c_0 + cn$, où c_0 représente le coût initial et c le coût de chaque nouvelle donnée. La fonction de perte à minimiser est alors donnée par l'expression

$$L_n(c) = r(\hat{\beta}_n, \beta_0) + C(n),$$

et si l'on fait le choix d'une erreur quadratique, on obtient

$$L_n(c) = \mathbb{E} \left(\hat{\beta}_n - \beta_0 \right)^2 + C(n).$$

D'après les hypothèses 2, $\sqrt{n}(\hat{\beta}_n - \beta_0)$ est asymptotiquement normal, et il existe $\sigma > 0$ tel que

$$\lim_{n \rightarrow \infty} n \mathbb{E}(\hat{\beta}_n - \beta_0)^2 = \sigma^2.$$

La fonction de perte est alors asymptotiquement équivalente à

$$\frac{\sigma^2}{n} + c_0 + cn,$$

lorsque $C(n) = c_0 + cn$. Il apparaît alors judicieux de considérer une règle d'arrêt sous la forme

$$N_c = \inf \left\{ n \geq n_0; n \geq \frac{\hat{\sigma}_n}{\sqrt{c}} \right\},$$

où $\hat{\sigma}_n$ est un estimateur de σ . Remarquons que cette règle d'arrêt est équivalente à la précédente.

Ici, il est possible de donner un sens à l'asymptotique en faisant tendre c vers 0. Effectivement, réduire le coût lié à l'ajout d'un individu et ne s'intéresser qu'à la qualité de l'observation revient à faire tendre la taille requise de l'échantillon vers l'infini.

Nous venons de voir que les préoccupations pour définir les règles d'arrêt sont multiples. Nous ne démontrerons des résultats sur les règles d'arrêt que dans le cas d'intervalles séquentiels. Ce sont ces mêmes règles d'arrêt qui feront l'objet de simulations dans le chapitre 7 pour le modèle à risques proportionnels.

Chapitre 2

Exemples de règles séquentielles d'arrêt et propriétés

Sommaire

2.1	Propriétés	45
2.2	Démonstration de la proposition 2.1.1	46
2.2.1	Existence de N_d	46
2.2.2	Divergence vers l'infini de N_d	46
2.2.3	Régularité	47
2.3	Généralisation de la variable d'arrêt	48
2.3.1	Divergence vers l'infini	49
2.3.2	Régularité	50

ans ce chapitre, nous revenons à la définition de la variable d'arrêt donnée dans le cadre d'intervalles séquentiels de confiance et nous montrons les différentes propriétés pour établir que la règle considérée est un élément de \mathcal{A} (voir chapitre 1, page 37). Par ailleurs, nous considérons les temps d'arrêt définis par un critère relatif à la valeur du paramètre β_0 et établissons un nouveau critère d'arrêt (nous verrons qu'il est possible d'introduire au sein de ce critère l'estimateur $\hat{\beta}_n$ de β_0) qui se révèle plus général que le précédent. Toutes les définitions et démonstrations qui suivent s'inscrivent dans le contexte simple où le paramètre de régression est de dimension 1. Le chapitre 6 est consacré au cas multidimensionnel.

Cela nous montre que l'ensemble \mathcal{A} est non vide, puisque l'un de ses représentants est exhibé. En outre, rappelons que les résultats du chapitre 3 sont valables quel que soit l'élément de \mathcal{A} considéré.

2.1 Propriétés

Nous nous mettons dans les conditions du paragraphe 1.4. Nous considérons $(\hat{\beta}_n)_{n \geq 1}$, une suite d'estimateurs d'un paramètre réel β_0 du modèle statistique où n est le nombre d'observations. Nous supposons que $(\hat{\beta}_n)_{n \geq 1}$ satisfait les hypothèses 2 du paragraphe 1.4.3 :

- (i) $\hat{\beta}_n \xrightarrow{P} \beta_0$, lorsque $n \rightarrow \infty$;
- (ii) $\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} \mathcal{N}(0, \sigma^2)$, lorsque $n \rightarrow \infty$, avec $\sigma > 0$;
- (iii) il existe une suite $(\hat{\sigma}_n^2)_{n \geq 1}$ d'estimateurs de σ^2 telle que :
 - (a) $\hat{\sigma}_n^2 \xrightarrow{P} \sigma^2$, lorsque $n \rightarrow +\infty$;
 - ou (b) $\hat{\sigma}_n^2 \xrightarrow{p.s.} \sigma^2$, lorsque $n \rightarrow +\infty$.

Nous considérons la règle d'arrêt définie, pour $d > 0, \alpha \in]0, 1[$, par :

$$N_d = \min \left\{ n \geq n_0; \frac{\hat{\sigma}_n^* u_\alpha}{\sqrt{n}} < d \right\},$$

où $\hat{\sigma}_n^* = \hat{\sigma}_n \vee \varepsilon_n$, $(\varepsilon_n)_{n \geq 1}$ est une suite de réels positifs tendant vers 0, et $\Phi(u_\alpha) = 1 - \alpha/2$, où Φ est la fonction de répartition d'une loi $\mathcal{N}(0, 1)$. Nous avons alors le résultat suivant.

Proposition 2.1.1. *Sous l'hypothèse (iii)-(a), N_d existe presque sûrement pour tout $d > 0$, et $(N_d)_{d > 0}$ diverge presque sûrement vers l'infini. De plus, si (iii)-(b) est satisfaite, $(N_d)_{d > 0}$ est un élément de \mathcal{A} (i.e. $(N_d)_{d > 0}$ est régulière).*

Notons que la plupart de ces propriétés fait l'objet de démonstration dans [GMS97], dans les cas de convergence forte à la fois pour l'estimateur de la variance et du paramètre de régression, dans un cadre non paramétrique. Dans un premier temps, considérant $\hat{\sigma}_n^{*2}$ comme un estimateur de la variance asymptotique σ^2 , nous vérifions sa convergence.

Lemme 2.1.1. $\hat{\sigma}_n^{*2} \xrightarrow{P} \sigma^2$, lorsque n tend vers l'infini.

Démonstration :

Considérons, pour $\varepsilon > 0$, l'événement $\{|\hat{\sigma}_n^* - \sigma| > \varepsilon\}$. Il s'écrit de la façon suivante :

$$\begin{aligned} \{|\hat{\sigma}_n^* - \sigma| > \varepsilon\} &= \{\max(\hat{\sigma}_n, \varepsilon_n) > \sigma + \varepsilon, \hat{\sigma}_n^* > \sigma\} \cup \{\max(\hat{\sigma}_n, \varepsilon_n) < \sigma - \varepsilon, \hat{\sigma}_n^* \leq \sigma\} \\ &= \{\hat{\sigma}_n \geq \varepsilon_n, \hat{\sigma}_n > \sigma + \varepsilon\} \cup \{\hat{\sigma}_n < \varepsilon_n, \varepsilon_n > \sigma + \varepsilon\} \\ &\cup \{\hat{\sigma}_n - \sigma < \varepsilon, \varepsilon_n < \sigma - \varepsilon\} \end{aligned}$$

La suite déterministe de réels strictement positifs $(\varepsilon_n)_{n \geq 1}$ converge vers 0, ce qui nous permet d'affirmer qu'on a, pour n assez grand et $0 < \varepsilon < \sigma/2$:

$$\{|\hat{\sigma}_n^* - \sigma| > \varepsilon\} \subset \{|\hat{\sigma}_n - \sigma| > \varepsilon\}, \quad (2.1)$$

ce qui permet d'obtenir le résultat, par l'hypothèse (iii)(a). \square

Lemme 2.1.2. *Sous l'hypothèse (iii)(b), $\hat{\sigma}_n^{*2} \xrightarrow{p.s.} \sigma^2$, lorsque n tend vers l'infini.*

Démonstration : Il suffit de considérer ici l'inclusion d'événements (2.1) établie dans la démonstration du lemme 2.1.1 et de remarquer que l'hypothèse (iii)(b) nous permet de conclure. \square

2.2 Démonstration de la proposition 2.1.1

Au cours de cette partie, nous nous attachons à démontrer l'existence et la régularité de la règle d'arrêt. Notons que notre démonstration de la régularité requiert de démontrer la divergence de la règle d'arrêt. Les propositions sont établies sous les hypothèses de la proposition 2.1.1.

2.2.1 Existence de N_d

Il existe une sous-suite $(N_{n(k)}^d)_{k \geq 1}$ convergeant presque sûrement vers σ , par (iii)-(a). Notons \mathcal{N}^c l'ensemble presque sûr sur lequel $(\hat{\sigma}_{n(k)}^*)_{k \geq 1}$ converge vers σ . Soit $\omega \in \mathcal{N}^c$, il existe $k_0 \in \mathbb{N}^*$ tel que, pour $k \geq k_0$,

$$0 < \hat{\sigma}_{n(k)}^* < 2\sigma.$$

Par conséquent, pour $m \in \{n(k); k \geq k_0\} \cap \left\{n \geq n_0 \vee \left(\frac{2u_\alpha \sigma}{d}\right)^2\right\}$, on a :

$$m \in \left\{n \geq n_0; \frac{u_\alpha \hat{\sigma}_n^*(\omega)}{\sqrt{n}} < d\right\};$$

ce dernier ensemble est donc non vide, ce qui assure l'existence de $N_d(\omega)$. Ceci étant vrai sur l'ensemble presque sûr \mathcal{N}^c , on obtient l'existence presque sûre de N_d . \square

2.2.2 Divergence vers l'infini de N_d

On remarque que pour tout $\omega \in \Omega$, $d \mapsto N_d(\omega)$ est une fonction à valeurs dans \mathbb{N} , croissante lorsque $d \rightarrow 0^+$, donc convergente dans $\overline{\mathbb{N}}$. Par conséquent, on doit démontrer que

$$\mathcal{V} = \{\omega \in \Omega; N_d(\omega) \rightarrow +\infty, d \rightarrow 0^+\}^c$$

est un ensemble négligeable. Si ω est un élément de \mathcal{V} , alors il existe d_0 tel que $d \mapsto N_d(\omega)$ est constante sur $]0, d_0]$. En effet, $(N_d(\omega))_{d > 0}$ est une suite d'entiers croissante lorsque d décroît, donc elle tend vers l'infini ou est constante à partir d'un certain rang. Par conséquent, pour $d \in]0, d_0]$, on a :

$$N_{d_0}(\omega) = N_d(\omega) > \frac{\hat{\sigma}_{N_{d_0}(\omega)}^*(\omega) u_\alpha}{d} \geq \frac{\varepsilon_{N_{d_0}(\omega)} u_\alpha}{d},$$

ce qui est impossible, pour d élément quelconque de $]0, d_0]$. On en déduit que :

$$\mathbb{P} \left(\lim_{d \rightarrow 0^+} N_d = +\infty \right) = \mathbb{P}(\mathcal{V}^c) = 1.$$

□

2.2.3 Régularité

Pour $d > 0$, on sait que $N_d(\omega)$ est élément de l'ensemble

$$\mathcal{V}_\omega = \left\{ n \geq n_0; \frac{u_\alpha \hat{\sigma}_n^*(\omega)}{\sqrt{n}} < d \right\},$$

et donc qu'on a simultanément

$$N_d(\omega) \geq n_0 \text{ et } N_d(\omega) > \frac{u_\alpha^2 \hat{\sigma}_{N_d(\omega)}^{2*}(\omega)}{d^2}. \quad (2.2)$$

Par ailleurs, $N_d(\omega)$ est défini comme le plus petit élément de l'ensemble \mathcal{V}_ω , ce qui signifie que $N_d(\omega) - 1$ n'est pas dans \mathcal{V}_ω , et donc que

$$\begin{aligned} N_d(\omega) - 1 &< n_0, \\ \text{ou } N_d(\omega) - 1 &\leq \left(\frac{u_\alpha \hat{\sigma}_{N_d(\omega)-1}^*(\omega)}{d} \right)^2. \end{aligned}$$

On peut rassembler ces deux cas en écrivant

$$N_d(\omega) \leq 1 + n_0 + \left(\frac{u_\alpha \hat{\sigma}_{N_d(\omega)-1}^*(\omega)}{d} \right)^2. \quad (2.3)$$

Il nous est possible d'obtenir des inégalités similaires pour n_d , soit :

$$\begin{aligned} n_d &\leq 1 + n_0 + \left(\frac{u_\alpha \sigma}{d} \right)^2, \\ n_d &\geq n_0, \\ n_d &\geq \frac{u_\alpha^2 \sigma^2}{d^2}. \end{aligned}$$

On en déduit que :

$$\frac{1}{(n_0 + 1) \left(\frac{d}{u_\alpha \hat{\sigma}_{N_d(\omega)}^*(\omega)} \right)^2 + \left(\frac{\sigma}{\hat{\sigma}_{N_d(\omega)}^*(\omega)} \right)^2} \leq \frac{N_d(\omega)}{n_d} \leq (n_0 + 1) \left(\frac{d}{u_\alpha \sigma} \right)^2 + \left(\frac{\hat{\sigma}_{N_d(\omega)-1}^*}{\sigma} \right)^2. \quad (2.4)$$

Remarque 2.2.1. *A ce niveau, l'hypothèse (iii)(a) sur $(\hat{\sigma}_n^2)_{n \geq 1}$ ne suffit pas à obtenir la convergence en probabilité de $(\hat{\sigma}_{N_d}^*)_{d > 0}$ vers σ , lorsque $d \rightarrow 0^+$.*

Alors, par 2.4, grâce à la divergence presque sûre de $(N_d)_{d > 0}$ et au lemme 2.1.2, on obtient le résultat souhaité en faisant tendre d vers 0. □

2.3 Généralisation de la variable d'arrêt

Il est question ici de généraliser au sein de la variable d'arrêt le paramètre d de sorte à ce qu'il ne corresponde pas uniquement à un réel strictement positif fixé par l'utilisateur, mais qu'il puisse dépendre de la valeur du paramètre de régression. Pour ce faire, on donnera dans un premier temps une nouvelle définition, dépendante de la taille n de l'échantillon, de la variable d'arrêt, dont on montrera ensuite les principales propriétés. Nous parlons ici de généralisation car cette nouvelle vision de l'amplitude maximale de l'intervalle de confiance englobe la précédente. Elle permet en outre d'insérer une décision relative à l'estimateur. n_0 et u_α représentent les mêmes quantités fixées au départ.

On peut constater expérimentalement que plus la quantité $|\beta_0|$ est grande (pour une variable explicative positive), plus il est nécessaire d'obtenir une grande taille d'échantillon pour un même d fixé. Cela peut s'expliquer par une plus grande hétérogénéité des données qui sont alors soumises aux variations de Z . C'est ainsi que nous sommes amenés à considérer une valeur butoir dépendante des données et plus particulièrement dépendant de la valeur de β_0 . Nous allons voir comment généraliser la règle d'arrêt et établir ses propriétés.

Définition 2.3.1. *Soit r , une fonction positive. Soit, pour n entier,*

$$d_n = d + \gamma r(\hat{\beta}_n),$$

avec $d > 0$ et $\gamma \geq 0$ donnés. On pose alors,

$$N_{d,\gamma} = \min \left\{ n \geq n_0; \frac{u_\alpha \hat{\sigma}_n^*}{\sqrt{n}} < d_n \right\},$$

$(d_n)_{n \geq 1}$ est une suite de réels strictement positifs.

Remarque 2.3.1. *Pour $\gamma = 0$, on retrouve la définition précédente.*

Proposition 2.3.1. *Sous les hypothèses 2 du paragraphe 1.4.3, pour tous $d > 0$ et $\gamma \geq 0$, $N_{d,\gamma}$ existe presque sûrement. Au cours de toute l'étude concernant la généralisation de la variable d'arrêt, on les supposera vérifiées.*

Démonstration :

Il suffit de reprendre la démonstration du paragraphe 2.2.1 en constatant que pour tout $\omega \in \mathcal{N}^c$, ensemble presque sûr, on a :

$$\left\{ n \geq n_0; \frac{u_\alpha \sigma_n^*(\omega)}{\sqrt{n}} < d \right\} \subset \left\{ n \geq n_0; \frac{u_\alpha \sigma_n^*(\omega)}{\sqrt{n}} < d + \gamma r(\hat{\beta}_n(\omega)) \right\}.$$

Comme l'ensemble de gauche est non vide, il en va de même pour celui de droite. \square

2.3.1 Divergence vers l'infini

On peut montrer un résultat de convergence d'une manière semblable à celle utilisée dans le paragraphe précédent en considérant par exemple des suites décroissantes de réels positifs dans la définition de d_n . On pose donc ici, pour p et n entiers, $d_{n,p} = d_p + \gamma_p r(\hat{\beta}_n)$, avec $(\gamma_p)_{p \geq 1}$ et $(d_p)_{p \geq 1}$, des suites décroissantes de réels positifs (respectivement strictement positifs) convergeant vers zéro quand p tend vers l'infini.

Proposition 2.3.2. *Soit r , une fonction positive et bornée. On suppose qu'on a, pour p entier, $d_{n,p} = d_p + \gamma_p r(\hat{\beta}_n)$, où $(d_p)_{p \geq 1}$ et $(\gamma_p)_{p \geq 1}$ sont des suites de réels respectivement strictement positifs et positifs, décroissantes, qui convergent vers 0, lorsque $p \rightarrow +\infty$. On note N_p , la règle définie par*

$$N_p = \min \left\{ n \geq n_0; \frac{u_\alpha \hat{\sigma}_n^*}{\sqrt{n}} < d_{n,p} \right\}.$$

Alors, on a :

$$N_p \xrightarrow{p.s.} +\infty,$$

lorsque $p \rightarrow +\infty$.

Remarque 2.3.2. *L'amplitude maximale dépend de la taille de l'échantillon, il nous est donc impossible d'indexer N par $d_{n,p}$, c'est pourquoi nous noterons N_p , où p représente l'indice des suites $(\gamma_p)_{p \geq 1}$ et $(d_p)_{p \geq 1}$.*

Remarque 2.3.3. *Pour n fixé, $(d_{n,p})_{p \geq 1}$ est une suite qui converge vers 0.*

Remarque 2.3.4. *Par décroissance de la suite $(d_{n,p})_{p \geq 1}$, à n fixé, $(N_p)_{p \geq 1}$ est presque sûrement croissante.*

Démonstration de la proposition 2.3.2 :

$(N_p)_{p \geq 1}$ est une suite d'entiers croissante, donc elle peut être stationnaire à partir d'un certain rang ou diverger vers l'infini. On reprend alors le raisonnement de la démonstration du paragraphe 2.2.2 en supposant cette suite stationnaire, et cela nous donne l'existence d'un entier positif p_0 tel que, pour tout entier p supérieur à p_0 , on ait

$$N_{p_0} = N_p > \frac{u_\alpha \varepsilon_{N_{p_0}}}{d_{N_{p_0}, p}},$$

ce qui est impossible, il suffit pour voir cela de faire tendre p vers l'infini et d'utiliser la remarque 2.3.3. \square

Corollaire 2.3.1. *La suite $(\hat{\sigma}_{N_p}^*)_{p \geq 1}$ converge presque sûrement vers σ .*

2.3.2 Régularité

Proposition 2.3.3. *Sous les hypothèses de la proposition 2.3.1, et si l'on suppose de plus que :*

- $(\hat{\beta}_n)_{n \geq 1}$ converge presque sûrement vers β_0 , lorsque $n \rightarrow +\infty$,
- r est une fonction continue,
- on a l'une des deux propriétés suivantes
 - (a) $d_{n,p} = d_p + \gamma_p r(\hat{\beta}_n)$, avec r strictement positive, et la suite $(\gamma_p/d_p)_{p \geq 1}$ converge vers 0,
 - (b) $d_{n,p} = d_p(1 + r(\hat{\beta}_n))$,

alors $(N_p)_{p \geq 1}$ est régulière.

Remarque 2.3.5. *Le cas (b) n'est pas une restriction de (a), puisque r n'est pas supposée strictement positive.*

Démonstration : Reprenons la démonstration du paragraphe 2.2.3. On définit ici, pour $p \geq 1$,

$$n_p = \min \left\{ n \geq n_0; n \geq \frac{u_\alpha \sigma}{d_{0,p}} \right\},$$

où $d_{0,p} = d_p + \gamma_p r(\beta_0)$, et l'on obtient l'encadrement suivant, inspiré par (2.4) :

$$\frac{1}{(1+n_0) \left(\frac{d_{N_p,p}}{u_\alpha \hat{\sigma}_{N_p}^*} \right)^2 + \left(\frac{\sigma}{\hat{\sigma}_{N_p}^*} \right)^2 \left(\frac{d_{N_p,p}}{d_{p,0}} \right)^2} \leq \frac{N_p}{n_p} \leq (1+n_0) \left(\frac{d_{p,0}}{u_\alpha \sigma} \right)^2 + \left(\frac{\hat{\sigma}_{N_p}^*}{\sigma} \right)^2 \left(\frac{d_{0,p}}{d_{N_p-1,p}} \right)^2.$$

On sait par le lemme 2.1.2 et la remarque 2.3.4 que $\hat{\sigma}_{N_p}^*$ converge presque sûrement vers σ . Par ailleurs, on sait que $d_{n_p,p}$ converge vers 0. Enfin, dans le cas (a), on a

$$\frac{d_{N_p,p}}{d_{0,p}} = \frac{1 + \gamma_p d_p^{-1} r(\hat{\beta}_{N_p})}{1 + \gamma_p d_p^{-1} r(\beta_0)} \xrightarrow{P} 1, \text{ lorsque } p \rightarrow +\infty,$$

et dans le cas (b), cela se traduit par

$$\frac{d_{N_p,p}}{d_{0,p}} = \frac{1 + r(\hat{\beta}_{N_p})}{1 + r(\beta_0)} \xrightarrow{P} 1, \text{ lorsque } p \rightarrow +\infty,$$

ce qui nous permet de conclure, dans les deux cas. □

Remarque 2.3.6. *Rappelons que la généralisation de la variable d'arrêt que nous venons d'introduire a pour but de construire des intervalles de confiance de longueur « contrôlée », mais relative à l'ordre de grandeur de β_0 . Le but est atteint, si par exemple, on prend, pour p fixé, $d_p = d + \gamma \left(|\hat{\beta}_n| \wedge A \right)$, avec $A > 0$.*

Découle de ce chapitre un exemple de règle d'arrêt appartenant à \mathcal{A} , établi dans un cadre semi-paramétrique. Nous considérons ensuite un élément quelconque de \mathcal{A} , noté $(N_d)_{d>0}$ et montrons les propriétés de l'estimateur séquentiel $\left(\hat{\beta}_{N_d} \right)_{d>0}$ dans le cadre semi-paramétrique. Cela nous permet de ne pas restreindre notre étude à la seule règle étudiée au cours du chapitre 2.

Chapitre 3

Théorèmes et propriétés des estimations séquentielles

Sommaire

3.1	Convergence presque sûre	53
3.1.1	Quelques résultats sur la convergence complète	53
3.1.2	Convergence forte	56
3.2	Normalité asymptotique	59

Ce chapitre vise à généraliser les théorèmes et résultats issus de [DG00] afin de les appliquer au cas du modèle semi-paramétrique. On supposera donc seulement ici que l'estimateur du paramètre de régression est issu d'une fonction d'estimation, et l'on s'attachera à montrer la convergence et la normalité de l'estimateur séquentiel d'arrêt, sous des conditions pré-établies. Elles feront l'objet de démonstrations pour le modèle de Cox dans le chapitre qui suit.

On considère ainsi le modèle semi-paramétrique présenté au cours du paragraphe 1.4. Soit $C_n(\beta)$, une fonction d'estimation obtenue pour n observations, où β_0 , la vraie valeur du paramètre de régression appartient à un ouvert borné de \mathbb{R}^p , noté \mathcal{B} . On suppose en outre que :

- $C_n \in C^2(\mathcal{B})$,
- C_n est concave dans un voisinage de β_0 ,
- $\hat{\beta}_n$ est un M -estimateur dans le sens où il est défini comme maximum d'une fonction. Ce n'est pas le sens habituel donné par exemple dans [Hub81] qui fait intervenir la notion de robustesse. On a ici :

$$\hat{\beta}_n = \operatorname{argmax}_{\beta \in \mathcal{B}} C_n(\beta).$$

On notera \mathcal{U}_n et \mathcal{I}_n , les dérivées première et seconde de C_n par rapport à β .

3.1 Convergence presque sûre

ous montrons ici que $\hat{\beta}_n$ est un estimateur fortement convergent de β_0 . À cette fin, nous allons établir la convergence complète de $(\hat{\beta}_n)_{n \geq 1}$ en nous inspirant d'une méthode donnée dans [DG00].

3.1.1 Quelques résultats sur la convergence complète

Nous allons donner dans un premier temps quelques résultats de stabilité sur les suites de variables aléatoires convergeant complètement, qui se révéleront utiles dans le cas du modèle de Cox. Ainsi, nous verrons qu'il suffit de nous ramener à montrer la convergence de processus élémentaires pour établir celle du processus d'intérêt, par intégration, somme ou même quotient.

Définition 3.1.1. Soit $(X_n)_{n \geq 1}$, une suite de variables aléatoires réelles. Soit X une variable aléatoire réelle. On dira que $(X_n)_{n \geq 1}$ converge complètement vers X et on notera $X_n \xrightarrow{c} X$, si :

$$\forall \varepsilon > 0, \sum_{n \geq 1} \mathbb{P}(|X_n - X| > \varepsilon) < \infty.$$

Remarque 3.1.1. La convergence complète implique la convergence presque sûre.

En effet : on considère une suite de variables aléatoires réelles $(X_n)_{n \geq 1}$ convergeant complètement vers X . On se ramène à une limite nulle en considérant la suite $(X_n - X)_{n \geq 1}$.

Si ε est un réel strictement positif, on a, pour n entier,

$$\mathbb{P}\left(\sup_{k \geq n} |X_k| > \varepsilon\right) \leq \sum_{k \geq n} \mathbb{P}(|X_k| > \varepsilon).$$

On utilise l'équivalence entre la convergence presque sûre de $(X_n)_{n \geq 1}$ vers 0 et la propriété

$$\forall \varepsilon > 0, \lim_{n \rightarrow \infty} \mathbb{P}\left(\sup_{k \geq n} |X_k| > \varepsilon\right) = 0,$$

pour vérifier l'inclusion souhaitée. □

Remarque 3.1.2. Il peut sembler naturel d'étudier la réciproque de la remarque précédente. Il est aisé de montrer, dans un cas simple, que les convergences presque sûre et complète ne sont pas équivalentes. Considérons en effet une suite $(X_n)_{n \geq 1}$ de variables aléatoires de Bernoulli, et notons, pour n entier positif, $p_n = \mathbb{P}(X_n = 1)$. Il est possible d'établir l'équivalence entre la convergence en probabilité de $(X_n)_{n \geq 1}$ vers 0 et la convergence vers 0 de la suite $(p_n)_{n \geq 1}$. En outre, la convergence de la série de terme général p_n nous donne la convergence presque sûre. Prenons maintenant X_n sous la forme suivante $X_n = \mathbb{1}_{[0, 1/n]}$ définies sur $([0, 1], \mathcal{B}([0, 1]), \mathbb{P})$,

où $\mathcal{B}([0, 1])$ est la tribu des boréliens sur $[0, 1]$ et \mathbb{P} la probabilité uniforme. Alors $(X_n)_{n \geq 1}$ converge en probabilité et presque sûrement, mais la convergence complète n'a pas lieu, puisque la série considérée ici n'est autre que la série harmonique. En outre, on peut remarquer que la suite de variables considérée dans ce contre-exemple est constituée de variables qui ne sont pas indépendantes, et que, dans le cas de variables indépendantes et toujours dans le cas de variables de Bernoulli, la convergence presque sûre équivaut à la convergence complète.

Remarque 3.1.3. Dans le cas de variables aléatoires indépendantes et identiquement distribuées, la convergence complète est effectivement équivalente à la convergence presque sûre [HR69].

Hsu et Robbins démontrent en outre que si $(Y_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes et identiquement distribuées, d'espérance nulle et de variance finie, alors leur moyenne empirique converge complètement vers 0. Cela relie ainsi la convergence complète à la loi des grands nombres.

Proposition 3.1.1.

Soient $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$, deux suites de variables aléatoires réelles convergeant complètement respectivement vers X et Y . Alors :

- (i) $(X_n + Y_n)_{n \geq 1}$ converge complètement vers $X + Y$,
- (ii) Si on suppose X et Y bornées, $(X_n Y_n)_{n \geq 1}$ converge complètement vers XY ,
- (iii) Si X est une variable presque sûrement minorée par une constante x strictement positive ($X > x > 0$ p.s.), $(1/X_n)_{n \geq 1}$ converge complètement vers $1/X$ (en prenant pour convention $1/0 = \infty$).

Démonstration :

- (i) Ramenons-nous au cas de limites nulles. Pour ce faire, il suffit de considérer $(X_n - X)_{n \geq 1}$ ainsi que $(Y_n - Y)_{n \geq 1}$. Pour ε un réel strictement positif, utilisons l'inclusion d'événements suivante

$$\{|X_n + Y_n| > \varepsilon\} \subset \left\{ |X_n| > \frac{\varepsilon}{2} \right\} \cup \left\{ |Y_n| > \frac{\varepsilon}{2} \right\}.$$

Cela nous permet de déduire l'inégalité correspondante en termes de probabilités. On peut ainsi conclure grâce à la convergence complète des suites $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$, et établir (i).

- (ii) Remarquons que :

$$X_n Y_n - XY = (X_n - X)(Y_n - Y) + X(Y_n - Y) + Y(X_n - X).$$

Par conséquent, il est possible de dresser le même type d'inclusions d'événements que lors de la démonstration précédente et de séparer ainsi trois termes. Pour obtenir le résultat, il nous suffit alors d'utiliser l'hypothèse de bornitude sur les limites des suites. Si l'on désigne par K , une constante strictement

positive pour laquelle $\max\{|X|, |Y|\} \leq K$ presque sûrement, on peut alors écrire

$$\begin{aligned} \mathbb{P}(|X_n Y_n - XY| > \varepsilon) &\leq \mathbb{P}(|X_n - X| > 1) + \mathbb{P}(|Y_n - Y| > \varepsilon/3) \\ &+ \mathbb{P}(|Y_n - Y| > \varepsilon/(3K)) \\ &+ \mathbb{P}(|X_n - X| > \varepsilon/(3K)). \end{aligned}$$

La décomposition du premier événement est obtenue par intersection avec $\{|Y_n - Y| < \varepsilon/3\}$. Par la convergence complète de $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$, on déduit la convergence souhaitée.

(iii) Soit \mathcal{N} , l'ensemble donné par $\mathcal{N} = \{\omega \in \Omega; X(\omega) \leq x\}$, il est négligeable et on a, pour tout réel ε strictement positif,

$$\begin{aligned} \mathbb{P}\left(\left|\frac{1}{X_n} - \frac{1}{X}\right| > \varepsilon\right) &\leq \mathbb{P}\left(\mathcal{N}^c, |X_n - X| \geq \frac{\varepsilon x^2}{2}\right) \\ &+ \mathbb{P}\left(\mathcal{N}^c, |X_n - X| \geq \varepsilon x |X_n|, |X_n| < \frac{x}{2}\right), \end{aligned}$$

en utilisant le fait que $\{|X_n| < x/2\} \subset \{|X_n - X| \geq x/2\}$, on obtient

$$\mathbb{P}\left(\left|\frac{1}{X_n} - \frac{1}{X}\right| > \varepsilon\right) \leq \mathbb{P}\left(|X_n - X| > \frac{\varepsilon x^2}{2}\right) + \mathbb{P}\left(|X_n - X| \geq \frac{x}{2}\right),$$

ce qui nous permet d'obtenir le résultat souhaité et d'achever la démonstration de la proposition. \square

Proposition 3.1.2.

Soit I un intervalle borné de \mathbb{R} . Soient $(f_n(t); t \in I)_{n \geq 1}$ et $(g_n(t); t \in I)_{n \geq 1}$, deux suites de processus aléatoires càdlàg telles que $(\sup_I |g_n - g|)_{n \geq 1}$ et $(\sup_I |f_n - f|)_{n \geq 1}$ convergent complètement vers 0. On suppose en outre que

1. $(f_n)_{n \geq 1}$ est à variations totales uniformément bornées, i.e. qu'il existe une constante B positive telle que pour tout $n \geq 1$, $\int_I |df| \leq B$.
2. g est continue sur I .

Alors la suite $(\int_I g_n df_n)_{n \geq 1}$ converge complètement vers $\int_I g df$.

Démonstration :

Le squelette de cette preuve est similaire à celui utilisé dans [BGY97]. Décomposons $\int_I g_n df_n - \int_I g df$ sous la forme

$$\int_I (g_n - g) df_n + \int_I g (df_n - df).$$

Comme g est supposée continue sur I , on peut trouver une suite $(g^{(l)})_{l \geq 1}$ de fonctions en escalier telles que, pour ε un réel strictement positif, il existe un entier l_0 pour lequel

$$\forall l \geq l_0, \sup_I |g^{(l)} - g| < \varepsilon.$$

Notons $a_0^l, \dots, a_{j_i}^l$, la subdivision adaptée à $g^{(l)}$, i.e. telle que l'on puisse écrire

$$g^{(l)} = \sum_{t=0}^{j_i-1} \alpha_t^l \mathbb{1}_{[a_t^l, a_{t+1}^l[}.$$

On obtient l'inégalité

$$\begin{aligned} \left| \int_I g_n df_n - \int_I g df \right| &\leq \left| \int_I (g_n - g) df_n \right| \\ &+ \left| \int_I (g - g^{(l)}) (df_n - df) \right| \\ &+ \left| \int_I g^{(l)} (df_n - df) \right|. \end{aligned}$$

En utilisant l'hypothèse 1., on voit que, pour $\eta > 0$,

$$\begin{aligned} \mathbb{P} \left(\left| \int_I g_n df_n - \int_I g df \right| > \eta \right) &\leq \mathbb{P} \left(B \sup_I |g_n - g| > \frac{\eta}{3} \right) \\ &+ \mathbb{P} \left(2B\varepsilon > \frac{\eta}{3} \right) \\ &+ \mathbb{P} \left(2 \sup_I |f_n - f| > \frac{\eta}{3H(\varepsilon)} \right), \end{aligned}$$

où $H(\varepsilon) = \sum_{t=0}^{j_i-1} |\alpha_t^l|$. Il suffit alors de prendre ε tel que $0 < \varepsilon < \eta/6B$, pour conclure. \square

3.1.2 Convergence forte

Nous disposons désormais d'outils concernant la convergence complète. Nous donnons tout d'abord un théorème permettant d'établir la convergence presque sûre de l'estimateur, inspiré par [DG00] où il figure dans un cadre paramétrique. Ce théorème est largement indépendant du modèle choisi, il suppose uniquement que l'estimateur du paramètre de régression est issu d'une fonction d'estimation C_n satisfaisant les conditions annoncées en début de chapitre.

Théorème 3.1.1.

On suppose que les quatre conditions suivantes sont vérifiées.

(C0) $\sqrt{n}U_n(\beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0))$, lorsque $n \rightarrow \infty$, où $\mathcal{I}(\beta_0) > 0$;

(C1) $|\mathcal{I}_n(\beta) - \mathbb{E}\mathcal{I}_n(\beta_0)|$ converge complètement vers 0 dans un voisinage de plus en plus petit de β_0 , ce qui signifie que l'on a, pour tout réel strictement positif ε ,

$$\exists \delta_\varepsilon > 0; \sum_{n \geq 1} \mathbb{P} \left(\sup_{\beta \in B(\beta_0; \delta_\varepsilon)} |\mathcal{I}_n(\beta) - \mathbb{E}\mathcal{I}_n(\beta_0)| > \varepsilon \right) < +\infty ;$$

$$(C2) \quad \mathbb{E}\mathcal{I}_n(\beta_0) \xrightarrow{n \rightarrow \infty} \mathcal{I}(\beta_0);$$

(C3) $(|U_n(\beta_0)|)_{n \geq 1}$ converge complètement vers 0, i.e.

$$\forall \varepsilon > 0, \sum_{n \geq 0} \mathbb{P}(|U_n(\beta_0)| > \varepsilon) < \infty.$$

Alors $\hat{\beta}_n$ tend vers β_0 pour la convergence complète, lorsque $n \rightarrow \infty$.

Démonstration :

On se place dans le cas où le paramètre de régression β_0 est un réel. Notre premier travail consiste à montrer que, pour δ un réel strictement positif, la propriété

$$\sum_{n \geq 1} \mathbb{P} \left(|U_n(\beta_0)| > \frac{\delta}{2} \inf_{\beta \in B(\beta_0; \delta)} \mathcal{I}_n(\beta) \right) < \infty \quad (3.1)$$

implique le résultat souhaité. Par ailleurs, il suffit de vérifier la condition définissant la convergence complète pour les valeurs de δ inférieures ou égales à un seuil strictement positif δ_ε , puisque l'inclusion d'événements élémentaires

$$\delta > \delta_\varepsilon \Rightarrow \left\{ |\hat{\beta}_n - \beta_0| > \delta \right\} \subset \left\{ |\hat{\beta}_n - \beta_0| > \delta_\varepsilon \right\}$$

nous permet de déduire le résultat pour les autres valeurs possibles de δ . Posons désormais $\varepsilon = C(\mathcal{I}(\beta_0))/2$, que l'on sait strictement positif par (C2), avec C une constante vérifiant $0 < C < 1$. Définissons l'ensemble A_n (pour n entier) par

$$A_n = \left\{ |U_n(\beta_0)| \leq \frac{\delta}{2} \inf_{\beta \in B(\beta_0; \delta)} \mathcal{I}_n(\beta) \right\}.$$

Par la formule de Taylor, on obtient, pour β élément de $\partial B(\beta_0, \delta)$, l'existence de $\tilde{\beta}$ dans le segment joignant β_0 et β tel que

$$C_n(\beta) - C_n(\beta_0) = u\delta U_n(\beta_0) - \frac{u^2\delta^2}{2} \mathcal{I}_n(\tilde{\beta}),$$

où $u = (\beta - \beta_0)/|\beta - \beta_0|$. Par définition de u , on sait que $|u| = 1$, et $u^2 = 1$, cela nous permet ainsi de déduire que l'on a, dans A_n :

$$|uU_n(\beta_0)| \leq \frac{u^2\delta}{2} \inf_{\beta \in B(\beta_0; \delta)} \mathcal{I}_n(\beta),$$

et d'obtenir ainsi :

$$A_n \subset \{\forall \beta \in \partial B(\beta_0; \delta), C_n(\beta) - C_n(\beta_0) \leq 0\} = D_n.$$

La concavité de la log-vraisemblance C_n en la variable β [AG82] permet de déduire l'inclusion

$$D_n \subset \left\{ \hat{\beta}_n \in B(\beta_0; \delta) \right\},$$

cela montre qu'il suffit d'établir (3.1).

Les probabilités $\mathbb{P}(A_n^c)$ étant difficiles à manier, on choisit d'intersecter les A_n^c avec les événements B_n définis, pour $C \in]0, 1[$, par

$$B_n = \left\{ \sup_{\beta \in B(\beta_0; \delta)} |\mathcal{I}_n(\beta) - \mathbb{E}(\mathcal{I}_n(\beta_0))| \leq C \mathbb{E}(\mathcal{I}_n(\beta_0)) \right\}.$$

Il est alors possible d'utiliser la décomposition suivante

$$\mathbb{P}(A_n^c) = \mathbb{P}(A_n^c \cap B_n) + \mathbb{P}(A_n^c \cap B_n^c),$$

et de se ramener ainsi à montrer que chaque terme du membre de droite est le terme général d'une série convergente.

On se place donc sur $A_n^c \cap B_n$, on voit que

$$(1 - C) \mathbb{E} \mathcal{I}_n(\beta_0) \leq \inf_{\beta \in B(\beta_0; \delta)} \mathcal{I}_n(\beta).$$

et que

$$|U_n(\beta_0)| > \left(\frac{1 - C}{2} \right) \delta \mathbb{E} \mathcal{I}_n(\beta_0).$$

On peut alors conclure pour ce premier terme, en utilisant (C2) et (C3).

Sur l'ensemble B_n^c , on établit l'inclusion, pour n assez grand

$$B_n^c \subset \left\{ \sup_{\beta \in B(\beta_0; \delta)} |\mathcal{I}_n(\beta) - \mathbb{E} \mathcal{I}_n(\beta_0)| > C \frac{\mathcal{I}(\beta_0)}{2} \right\},$$

il nous suffit alors d'utiliser la condition (C1) pour conclure. \square

Corollaire 3.1.1. *Pour toute suite régulière $(T_n)_{n \geq 1}$, $(\hat{\beta}_{T_n})_{n \geq 1}$ converge en probabilité vers β_0 .*

Démonstration : Grâce à la régularité de $(T_n)_{n \geq 1}$, on sait qu'il existe une suite entière croissante $(t_n)_{n \geq 1}$, telle que $(T_n/t_n)_{n \geq 1}$ converge en probabilité vers 1. Considérons la suite croissante $(t_n^*)_{n \geq 1}$ définie par :

$$t_n^* = \inf \{ (1 - \eta) t_k; k \geq n \},$$

où $\eta \in]0, 1[$. Pour $n \geq 1$ et $\varepsilon > 0$, on écrit :

$$\begin{aligned} & \mathbb{P} \left(\left| \hat{\beta}_{T_n} - \beta_0 \right| > \varepsilon \right) \\ & \leq \mathbb{P} (T_n < t_n^*) + \sum_{k \geq t_n^*} \mathbb{P} \left(\left| \hat{\beta}_k - \beta_0 \right| > \varepsilon \right). \end{aligned}$$

Le premier terme converge vers 0 par régularité de $(T_n)_{n \geq 1}$ et le second est le reste d'une série convergente. \square

3.2 Normalité asymptotique

Il est maintenant possible d'établir un théorème permettant d'obtenir les résultats de normalité asymptotique désirés. Notons qu'une seule hypothèse est ajoutée, elle correspond à une condition de type Anscombe. Nous reviendrons sur cette hypothèse dans le chapitre suivant, lors de sa vérification dans le cadre du modèle à risques proportionnels de Cox. Ce théorème généralise celui de [DG00] dans le sens où il est adapté ici aux modèles semi-paramétriques. Il est tout d'abord démontré dans le cas de paramètres réels et fera l'objet d'une généralisation au cas multidimensionnel dans le chapitre 6.

Théorème 3.2.1.

Supposons vérifiées les conditions (C0) à (C3) du théorème 3.1.1, et qu'en outre, on ait

(C4) Pour toute suite régulière $(T_n)_{n \geq 1}$ de variables aléatoires entières,

$$\sqrt{T_n} U_{T_n}(\beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)), \text{ lorsque } n \rightarrow \infty.$$

Alors

$$\sqrt{T_n}(\hat{\beta}_{T_n} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)^{-1}), \text{ lorsque } n \rightarrow \infty.$$

Remarque 3.2.1. Il est suffisant, pour un modèle donné vérifiant les conditions (C0) à (C4), de montrer la régularité d'une règle d'arrêt pour établir ses propriétés asymptotiques.

Lemme 3.2.1. Sous les hypothèses du théorème 3.2.1, si η_{T_n} est inclus dans le segment qui joint β_0 et $\hat{\beta}_{T_n}$, alors :

- (i) $\mathcal{I}_n(\eta_n) \xrightarrow{c} \mathcal{I}(\beta_0)$, lorsque n tend vers l'infini,
- (ii) $\mathcal{I}_{T_n}(\eta_{T_n}) \xrightarrow{P} \mathcal{I}(\beta_0)$, lorsque n tend vers l'infini.

Démonstration :

Montrons dans un premier temps que le résultat est vérifié pour la suite $(\eta_n)_{n \geq 1}$. Pour ε et $\delta = \delta_{\varepsilon/2}$ strictement positifs, on a

$$\{|\mathcal{I}_n(\eta_n) - \mathcal{I}(\beta_0)| > \varepsilon\} \subset \left\{ \sup_{\beta \in B(\beta_0, \delta)} |\mathcal{I}_n(\beta) - \mathbb{E}\mathcal{I}_n(\beta_0)| > \frac{\varepsilon}{2} \right\} \quad (\text{B1})$$

$$\cup \left\{ |\mathbb{E}\mathcal{I}_n(\beta_0) - \mathcal{I}(\beta_0)| > \frac{\varepsilon}{2} \right\} \quad (\text{B2})$$

$$\cup \left\{ \hat{\beta}_n \subset B^c(\beta_0, \delta) \right\}. \quad (\text{B3})$$

(B1) est le terme général d'une série convergente par (C1), il suffit en outre d'appliquer (C2) pour déduire un résultat analogue pour (B2). C'est ensuite le théorème 3.1.1 qui implique que (B3) est sommable, et établit finalement

$$\sum_{n \geq 1} \mathbb{P}(|\mathcal{I}_n(\eta_n) - \mathcal{I}(\beta_0)| > \varepsilon) < +\infty. \quad (3.3)$$

Il vient que $(\mathcal{I}_n(\eta_n))_{n \geq 1}$ converge presque sûrement vers $\mathcal{I}(\beta_0)$. Montrons maintenant (ii).

Grâce à la régularité de $(T_n)_{n \geq 1}$, on sait qu'il existe $(t_n)_{n \geq 1}$, telle que T_n/t_n converge en probabilité vers 1.

C'est par le biais de la suite croissante $(t_n^*)_{n \geq 1}$ (définie au cours de la démonstration du corollaire 3.1.1) qu'il sera possible d'utiliser les hypothèses de régularité sur la suite de variables d'arrêt. On rappelle que

$$t_n^* = \inf \{(1 - \eta)t_k; k \geq n\}.$$

L'inclusion correspondante des événements permet de déduire

$$\begin{aligned} & \mathbb{P} (|\mathcal{I}_{T_n}(\eta_{T_n}) - \mathcal{I}(\beta_0)| > \varepsilon) \\ & \leq \mathbb{P} (T_n < t_n^*) + \sum_{k \geq t_n^*} \mathbb{P} (|\mathcal{I}_k(\eta_k) - \mathcal{I}(\beta_0)| > \varepsilon). \end{aligned}$$

On peut alors conclure car le premier terme converge vers 0 et le second est le reste de la somme (3.3). \square

Corollaire 3.2.1. *Sous les hypothèses du théorème 3.2.1, on a :*

$$\mathcal{I}_n(\hat{\beta}_n)^{-1} \xrightarrow{c} \mathcal{I}(\beta_0)^{-1}, n \rightarrow +\infty.$$

Démonstration : C'est une conséquence du lemme 3.2.1 ci-dessus et de la proposition 3.1.1 (ii) prise avec (C0) qui indique que $\mathcal{I}(\beta_0) > 0$. \square

Démonstration du théorème 3.2.1 :

Par définition de U_n et en appliquant la formule de Taylor en β_0 , on obtient

$$\sqrt{T_n} (\hat{\beta}_{T_n} - \beta_0) = -\sqrt{T_n} U_{T_n}(\beta_0) \mathcal{I}_{T_n}^{-1}(\eta_{T_n}),$$

où η_{T_n} se situe sur le segment reliant β_0 et $\hat{\beta}_{T_n}$.

Par (C4), on sait que $\sqrt{T_n} U_{T_n}(\beta_0)$ est asymptotiquement normal. Par ailleurs, le lemme précédent ainsi que (C2) nous permettent, en utilisant le théorème de Slutsky ainsi que la normalité asymptotique de $\sqrt{T_n} U_{T_n}(\beta_0)$, de montrer que $\sqrt{T_n}(\hat{\beta}_{T_n} - \beta_0)$ est asymptotiquement normal, ce qui achève la démonstration de ce théorème. \square

Remarque 3.2.2. *Notons qu'au cours des deux démonstrations précédentes, il est possible de se ramener à un cadre multidimensionnel par le simple fait de remplacer les valeurs absolues par une norme sur \mathbb{R}^p ($p \geq 2$). Il ne sera donc aucunement nécessaire de reprendre cette démonstration dans le cas $p > 1$.*

Il nous reste désormais à vérifier ces conditions pour le modèle à risques proportionnels de Cox en utilisant les propriétés de stabilité de la convergence complète. C'est l'objet du chapitre suivant.

Chapitre 4

Analyse séquentielle pour le modèle de Cox

Sommaire

4.1	Convergence	63
4.1.1	Hypothèses et remarques	63
4.1.2	Convergence complète : démonstration de (C1) et (C3)	65
4.2	Normalité asymptotique : preuve de (C4)	73
4.2.1	L'équivalence complète et asymptotique	74
4.2.2	Démonstration de (C4)	75

Nous voulons désormais établir des propriétés de l'estimateur du paramètre de régression pour le modèle de Cox dans un cadre séquentiel. Plus précisément, nous nous situons maintenant dans le cadre du paragraphe 1.3, c'est-à-dire que nous observons pour n individus des durées de vie censurées ainsi que les indicateurs de censure. Au cours de ce chapitre, nous reprenons les notations du paragraphe 1.3. Les données sont sujettes à la censure à droite par le biais d'une variable C de censure (les notations des chapitres précédents seront utilisées). La règle séquentielle d'arrêt est fixée, et les quelques propriétés s'y rapportant sont supposées vérifiées (cela signifie entre autres que la règle d'arrêt considérée est régulière). Dans cette partie, nous nous attacherons à démontrer des conditions suffisantes pour établir la convergence, puis la normalité asymptotique de l'estimateur de β_0 pour une taille d'échantillon aléatoire T_n , donnée par la règle d'arrêt. Ce travail fait l'objet d'une prépublication [BB03] soumise pour publication. Des deux propriétés à montrer se dessinent naturellement deux parties de ce chapitre ; les techniques utilisées diffèrent également, bien qu'elles se ramènent toutes deux à établir des inégalités exponentielles. Du fait de la relative complexité des démonstrations et des notations s'y rattachant, on considère désormais que le paramètre de régression est réel, ce qui signifie qu'une seule variable explicative est repertoriée dans l'étude. Nous verrons dans le chapitre 6 que la généralisation à un espace de dimension p est possible.

4.1 Convergence

Comme nous l'avons vu au cours des chapitres précédents, il paraît peu judicieux de tenter d'établir des propriétés sur les estimateurs séquentiels, sans bénéficier d'hypothèses sur les estimateurs considérés. Ainsi, il se révèle naturel de supposer la convergence, la normalité asymptotique et la convergence de l'éventuel estimateur de la variance asymptotique. Pour le modèle à risques proportionnels de Cox, on considère des hypothèses légèrement plus faibles que celles d'[AG82] (répertoriées page 31), et nous verrons qu'elles nous permettent d'établir un résultat de convergence normale de l'estimateur du paramètre de régression. Notons toutefois que le modèle considéré ici est moins général que celui qui est traité dans [AG82]. En effet, on suppose ici que chaque individu fait l'expérience de \mathcal{E} (l'événement d'intérêt) au plus une fois au cours de sa vie.

4.1.1 Hypothèses et remarques

Théorème 4.1.1.

On considère les hypothèses suivantes.

H1 - $\int_0^\tau \lambda_0(t)dt$ est fini ;

H2bis - Il existe un voisinage \mathcal{B} de β_0 , des fonctions $s^{(0)}$, $s^{(1)}$ et $s^{(2)}$ définies sur $\mathcal{B} \times [0, \tau]$ telles que, pour $j = 0, 1, 2$,

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} |\mathbb{E}S_n^{(j)}(\beta, t) - s^{(j)}(\beta, t)| \rightarrow 0;$$

H3 - En conservant les notations précédentes, on définit

$$e = \frac{s^{(1)}}{s^{(0)}} \text{ et } v = \frac{s^{(2)}}{s^{(0)}} - e^{\otimes 2}.$$

Pour β dans \mathcal{B} , $t \in [0, \tau]$,

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s^{(0)}(\beta, t), s^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta^2} s^{(0)}(\beta, t),$$

$s^{(0)}(\cdot, t)$, $s^{(1)}(\cdot, t)$, et $s^{(2)}(\cdot, t)$ sont des fonctions continues sur \mathcal{B} , uniformément en $t \in [0, \tau]$; $s^{(0)}$ est strictement positive sur $\mathcal{B} \times [0, \tau]$, et

$$\mathcal{I}(\beta_0) = \int_0^\tau v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt$$

est strictement positif ;

H4 - $\{Z_i(t); t \in [0, \tau]\}$, pour i entier naturel, est à variations bornées, uniformément en $i \geq 1$, i.e.

$$\forall i \geq 1, \int_0^\tau |Z_i(ds)| \leq B < \infty,$$

où B désigne une constante positive.

Alors, les conditions (C0) à (C4) du chapitre 3 sont vérifiées.

Corollaire 4.1.1. *Sous (H1)-(H4), $\hat{\beta}_n$ converge complètement vers β_0 . En outre, pour toute suite régulière $(T_n)_{n \geq 1}$ de variables d'arrêt, on a :*

- (i) $\hat{\sigma}_n^2 = \mathcal{I}_n(\hat{\beta}_n, \tau)^{-1} \xrightarrow{c} \mathcal{I}(\beta_0)^{-1}$, $n \rightarrow +\infty$;
- (ii) $\sqrt{T_n}(\hat{\beta}_{T_n} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)^{-1})$, $n \rightarrow +\infty$.

Démonstration : D'après le théorème 4.1.1, (C0)-(C4) sont satisfaites, les résultats (i) et (ii) sont donc des conséquences du corollaire 3.2.1 et du théorème 3.2.1. \square

Remarque 4.1.1. *Les résultats du corollaire 4.1.1 ci-dessus permettent d'affirmer que la règle d'arrêt $(N_d)_{d>0}$ définie dans le paragraphe 2.1.1 est un élément de \mathcal{A} . C'est aussi le cas pour la règle d'arrêt résultant du critère d'arrêt relatif à $\hat{\beta}_n$, puisque l'on a la convergence presque sûre de $\hat{\beta}_n$ vers β_0 .*

Remarque 4.1.2 (Discussion sur les hypothèses (H1)-(H4)).

(H1) *est analogue à l'hypothèse émise par [AG82]. Considérer une étude sur un intervalle de temps du type $[0, \tau]$ semble restrictif; néanmoins Andersen et Gill [AG82] ont montré, dans le cas de variables indépendantes et identiquement distribuées (avec des variables explicatives bornées, et certaines conditions sur N et Y), que la contribution à la matrice d'information de Fisher des données sur $[\tau, +\infty[$ peut être rendue arbitrairement petite, uniformément en n , pour une valeur de τ assez élevée. Pour plus de détails, on se reportera à [AG82, théorème 4.2 p.1112], ainsi qu'aux discussions qui s'y rapportent.*

(H2bis) *Elle se présente sous une forme moins restrictive que l'hypothèse donnée dans [AG82]. Nous verrons cependant que la condition (H2) est automatiquement vérifiée, sous les hypothèses (H1), (H2bis), (H3) et (H4).*

(H4) *Cette condition est par exemple utilisée dans [BGY97] (il s'agit de la Condition 1) pour l'étude du modèle de Cox à entrées retardées. Elle se révèle essentielle dans la plupart de nos preuves, et l'indépendance de la borne par rapport à l'individu considéré est elle aussi importante. Enfin, la plupart des applications satisfont cette condition, qui est donc peu restrictive.*

Dans la suite de ce chapitre, nous montrons les propriétés (C0)-(C4) une à une. Dans un premier temps, nous verrons que certaines de ces conditions se vérifient de façon immédiate ou résultent d'hypothèses classiques.

Ainsi, (C0) est vérifiée, car elle résulte de la démonstration d'Andersen et Gill [AG82, théorème 3.2 p.1106]. Cependant, utiliser le résultat montré par Andersen et Gill nécessite d'obtenir des hypothèses qui impliquent (H1) à (H3). Il suffit pour cela de considérer le lemme 4.1.1 et de constater que sa démonstration n'est nullement assujettie à la condition (C0).

(C2) est une hypothèse classique. En effet, dans le cas de variables indépendantes et identiquement distribuées, il est possible d'appliquer une loi des grandes nombres

pour les $S_n^{(i)}$ et de constater que leur limite n'est autre que $\mathbb{E}Y(t)Z(t)^i \exp(\beta Z(t))$, où Z , et Y sont des représentants de mêmes lois que, respectivement Z_1 et Y_1 (voir [ABGK93, AG82]).

(C1), (C3) et (C4) nécessitent davantage de calculs. Dans un premier temps, nous obtenons la convergence complète de l'estimateur du paramètre de régression par le théorème 3.1.1 en établissant (C1) ainsi que (C3). Suivra alors la vérification que (C4) qui apportera, par le théorème 3.2.1, des résultats de normalité asymptotique pour les estimateurs séquentiels.

Dans toute la suite, le paramètre β est omis, lorsqu'il est égal à sa vraie valeur β_0 . La partie suivante est consacrée à la démonstration des conditions (C1) et (C3) du théorème 4.1.1.

4.1.2 Convergence complète : démonstration de (C1) et (C3)

Les techniques issues de [Pol90, vdVW96] sont utilisées, les principaux résultats et notations sont répertoriés dans l'Annexe A, ainsi que des rappels de définitions. Remarquons ici que les premiers résultats énoncés sont plus généraux que nécessaire pour établir uniquement (C3). Néanmoins la démonstration de (C1) les requiert. Par soucis de concision et de clarté, nous les énonçons directement sous une forme plus générale. Les notations de cette première démonstration correspondent en majeure partie avec celles de [Pol90].

Notation 1.

Considérons un tableau triangulaire de processus

$$\mathcal{F}_{n\omega} = \{f_{ni}(\omega, t); t \in [0, \tau], i \in \{1, \dots, k_n\}\}.$$

En pratique, on aura le plus souvent $k_n = n$. Nous nous intéressons à la convergence complète de

$$\Delta_{n\omega} = \sup_{t \in [0, \tau]} |S_n(\omega, t) - \mathbb{E}S_n(\cdot, t)|,$$

où $S_n(\omega, t) = \sum_{i=1}^{k_n} f_{ni}(\omega, t)$. On notera $\delta_{n\omega} = \sup_{f \in \mathcal{F}_{n\omega}} \|f\|_2$, où $\|\cdot\|_2$ désigne la norme euclidienne.

Notre premier objectif est de montrer la convergence complète de la fonction score, notée U_n . Nous remarquons que cette fonction s'écrit sous forme de somme de processus. Il s'avère alors utile de décomposer U_n en termes simples dont on sait montrer la convergence complète. À cette fin, nous avons besoin des lemmes qui suivent (leurs démonstrations seront données après la preuve de (C3), qui suit immédiatement).

Lemme 4.1.1. *Pour m dans $\{0, 1, 2\}$, $\sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} |S_n^{(m)}(\beta, t) - s^{(m)}(\beta, t)|$ converge complètement vers 0, où \mathcal{B}' désigne un voisinage borné de β_0 .*

Lemme 4.1.2 (corollaire du lemme 4.1.1). *Il existe un voisinage borné \mathcal{B}' de β_0 tel que*

$$\sup_{(\beta,t) \in \mathcal{B}' \times [0,\tau]} \left| \frac{S_n^{(1)}(\beta,t)}{S_n^{(0)}(\beta,t)} - \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)} \right|$$

converge complètement vers 0.

Lemme 4.1.3. *La suite $\left(\sup_{t \in [0,\tau]} \left| \frac{1}{n} \sum_{i=1}^n \int_0^t Z_i(s) dM_i(s) \right| \right)_{n \geq 1}$ converge complètement vers 0.*

Lemme 4.1.4. $\left(\sup_{t \in [0,\tau]} \left| \frac{1}{n} \sum_{i=1}^n M_i(t) \right| \right)_{n \geq 1}$ *converge complètement vers 0. De plus, $(\frac{1}{n} \sum_{i=1}^n M_i(t))_{n \geq 1}$ est une suite de processus à variations totales bornées.*

Démonstration de (C3) :

Pour $t \in [0, \tau]$, on écrit

$$|U_n(\beta_0, t)| \leq \left| \frac{1}{n} \sum_{i=1}^n \int_0^t Z_i(s) dM_i(s) \right| \quad (4.1)$$

$$+ \left| \frac{1}{n} \sum_{i=1}^n \int_0^t (E_n(\beta_0, s) - e(\beta_0, s)) dM_i(s) \right|, \quad (4.2)$$

$$+ \left| \frac{1}{n} \sum_{i=1}^n \int_0^t e(\beta_0, s) dM_i(s) \right|, \quad (4.3)$$

où $M_i(t) = N_i(t) - \int_0^t \lambda_0(s) Y_i(s) \exp(\beta_0 Z_i(s)) ds$ est une martingale. La convergence complète de (4.1) découle du lemme 4.1.3, celle du terme (4.2) provient du lemme 4.1.2 qui donne la convergence de $\sup_{t \in [0,\tau]} |E_n(\beta_0, t) - e(\beta_0, t)|$ vers 0 ainsi que du lemme 4.1.4 où il est établi que les variations totales de $n^{-1} \sum_{i=1}^n M_i(t)$ sont bornées. Il reste à montrer que le terme (4.3) converge complètement vers 0. On souhaite utiliser le corollaire B.1.1 pour montrer la convergence complète du dernier terme. À cette fin, on remarque que $M^{(n)}(t) = n^{-1} \sum_{i=1}^n \int_0^t e(\beta_0, s) dM_i(s)$ est une martingale locale, localement uniformément intégrable, nulle en 0, qui vérifie

$$\|\Delta M^{(n)}\|_0^\tau \leq \frac{c}{n}, \text{ p.s.}$$

En effet,

$$M^{(n)}(t) = \frac{1}{n} \sum_{i=1}^n \int_0^t e(\beta_0, s) dN_i(s) - \int_0^t e(\beta_0, s) S_n^{(0)}(\beta_0, s) \lambda_0(s) ds,$$

et comme le second terme du membre de droite de l'inégalité ci-dessus est continu, il vient :

$$\Delta M^{(n)}(t) = \frac{1}{n} \sum_{i=1}^n e(\beta_0, X_i) \Delta N_i(t).$$

Or, $e(\beta_0, \cdot)$ est bornée et $\Delta\bar{N}(t) \leq 1$ presque sûrement, d'où :

$$\|\Delta M^{(n)}\|_0^\tau \leq \frac{c}{n},$$

où c est une constante. Donc la première hypothèse du corollaire est vérifiée. Pour la seconde, on remarque que, pour $t \in [0, \tau]$,

$$\langle M^{(n)} \rangle (t) = \frac{1}{n^2} \sum_{i=1}^n \int_0^t e(\beta_0, s)^2 \lambda_0(s)^2 Y_i(s) ds$$

est majorée par d/n (où d est une constante positive), et donc que la quantité

$$\mathbb{P} (\langle M^{(n)} \rangle (t) \geq n^{-1/4})$$

est nulle à partir d'un certain rang, ce qui nous permet d'appliquer le corollaire et de déduire que $(\|M^{(n)}\|_0^\tau)_{n \geq 1}$ converge complètement vers 0. Cela achève la démonstration de (C3). \square

Lemme 4.1.5 (*prélude à la démonstration du lemme 4.1.1*).

La famille de processus \mathcal{H} donnée par

$$\mathcal{H} = \{(\exp(\beta Z_1(t)), \dots, \exp(\beta Z_n(t))); t \in [0, \tau], \beta \in \mathcal{B}\},$$

où \mathcal{B} est un voisinage borné de β_0 , est euclidienne.

Démonstration du lemme 4.1.5 :

Il nous est impossible de montrer que \mathcal{H} est euclidienne par un résultat de stabilité issu par exemple de [Pol90], puisque la famille \mathcal{H} ne se présente pas directement sous forme de produit terme à terme d'éléments de familles de processus contrôlables; et les résultats de contrôlabilité suite à la monotonie de fonctions ne s'appliquent pas ici. Les paramètres de la famille considérée possèdent des caractéristiques fortes comme la monotonie par rapport à chacune des deux variables t et β et une certaine indépendance par rapport à son enveloppe $\mathbb{H} = (H_1, \dots, H_n)$, puisque les H_i ne dépendent pas de i . C'est ainsi qu'il paraît naturel d'établir un résultat de contrôlabilité de base, c'est-à-dire qui fasse intervenir les concepts définissant cette notion. Un second ingrédient pour le contrôle de ce processus est apporté par la convexité de l'exponentielle *via* le résultat suivant.

Lemme 4.1.6. (*cf [Döh00]*) *Soit \mathcal{F} , un sous-ensemble de \mathbb{R}^n . Pour tout $f \in \mathcal{F}$: $\|f\|_2 \leq K$ et φ fonction définie sur la boule centrée de rayon K pour la distance euclidienne, à valeurs dans \mathbb{R}^n est une fonction $Lip(\varphi)$ -lipschitzienne. Alors*

$$D(\varepsilon, \varphi \circ \mathcal{F}) \leq D\left(\frac{\varepsilon}{Lip(\varphi)}, \mathcal{F}\right),$$

où $\varphi \circ \mathcal{F} = \{\varphi(f); f \in \mathcal{F}\}$.

Pour $\gamma = (\gamma_1, \dots, \gamma_n)$ des poids positifs fixés, on souhaite établir l'existence de constantes positives A et W telles que

$$D(\varepsilon \|\gamma \odot \mathbb{H}\|_2, \gamma \odot \mathcal{H}) \leq A \left(\frac{1}{\varepsilon} \right)^W.$$

Notre premier travail consiste alors à s'affranchir de la présence de poids positifs car ils interdisent l'utilisation du lemme 4.1.6. Nous verrons que les particularités de la famille de processus considérée le permettent.

Dans le but de définir \mathcal{H} comme résultante d'une composition terme à terme par la fonction exponentielle, introduisons \mathcal{G} le sous-ensemble de \mathbb{R}^n défini par

$$\mathcal{G} = \{(\beta Z_1(t), \dots, \beta Z_n(t)); t \in [0, \tau], \beta \in \mathcal{B}\}.$$

On considère des poids positifs fixés $\gamma \in \mathbb{R}^n$.

Pour $g \in \mathcal{G}$, on remarque que $\|g\|_2 \leq \|(B \times C(\mathcal{B}), \dots, B \times C(\mathcal{B}))\|_2$, où $C(\mathcal{B}) = \sup\{\|\beta\|; \beta \in \mathcal{B}\}$. On prend donc pour \mathcal{G} l'enveloppe $\mathbb{G} = (G, \dots, G)$ avec $G = B \times C(\mathcal{B})$ qui est indépendante de la direction. Dans ce cas, on a l'égalité qui suit

$$D(\varepsilon \|\gamma \odot \mathbb{H}\|_2, \gamma \odot \mathcal{H}) = D(\varepsilon \|\mathbb{H}\|_2, \alpha \odot \mathcal{H}),$$

où α est un poids positif de norme 1 (pour $1 \leq i \leq n$, $\alpha_i = \gamma_i / \|\gamma\|_2$), en remarquant que $\mathbb{H} = (\exp G, \dots, \exp G) = \exp \mathbb{G}$ est une enveloppe pour \mathcal{H} .

Puisque la fonction λ_α de \mathbb{R}^n vers \mathbb{R}^n définie par $\lambda_\alpha(x) = \alpha \odot x$ est une contraction (au sens large), on obtient, en utilisant [Pol90, p.24], que :

$$D(\varepsilon \|\mathbb{H}\|_2, \alpha \odot \mathcal{H}) \leq D(\varepsilon \|\mathbb{H}\|_2, \mathcal{H}).$$

En utilisant le lemme 4.1.6 on obtient que

$$\begin{aligned} D(\varepsilon \|\mathbb{H}\|_2, \mathcal{H}) &\leq D(\varepsilon \|\exp(\mathbb{G})\|_2, \exp \circ \mathcal{G}) \\ &\leq D(\varepsilon e^{-K} \|\exp(\mathbb{G})\|_2, \mathcal{G}), \end{aligned}$$

avec $K = B \times C(\mathcal{B})$.

\mathcal{G} est un produit terme à terme de $\{(\beta, \dots, \beta); \beta \in \mathcal{B}\}$ dont la pseudo-dimension est 1 et de $\{(Z_1(t), \dots, Z_n(t)); t \in [0, \tau]\}$ qui peut être exprimée comme différence de fonctions positives croissantes et donc euclidiennes. On déduit que \mathcal{G} est euclidienne avec le lemme A.3.3.

On obtient alors l'existence de constantes A et W telles que, pour $0 < \kappa \leq 1$,

$$D(\kappa \|\mathbb{G}\|_2, \mathcal{G}) \leq A \left(\frac{1}{\kappa} \right)^W.$$

Il reste ensuite à remarquer que

$$D(\varepsilon e^{-K} \|\exp \mathbb{G}\|_2, \mathcal{G}) \leq D(\varepsilon e^{-K} \|\mathbb{G}\|_2, \mathcal{G}),$$

en utilisant le fait que pour $x > 0$, $e^x \geq x$, et en prenant $\kappa = \varepsilon e^{-K} \in]0, 1]$ pour déduire l'inégalité suivante :

$$D(\varepsilon \|\gamma \odot \mathbb{H}\|_2, \gamma \odot \mathcal{H}) \leq (Ae^{KW}) \times \left(\frac{1}{\varepsilon}\right)^W,$$

et donc \mathcal{H} est euclidienne. \square

Il nous reste cependant à vérifier le lemme 4.1.6.

Preuve du lemme 4.1.6 : En utilisant la définition de la constante de Lipschitz, on constate aisément que pour x et y éléments de \mathcal{F} , on a

$$\|\varphi(x) - \varphi(y)\| > \varepsilon \Rightarrow \|x - y\| > \frac{\varepsilon}{\text{Lip}(\varphi)},$$

et l'on déduit alors le résultat par définition de D . \square

Démonstration du lemme 4.1.1 :

Remarquons d'abord qu'il suffit de vérifier le résultat pour $(m \in \{0, 1, 2\})$

$$\sup_{\beta \in \mathcal{B}'} \sup_{t \in [0, \tau]} |S_n^{(m)}(\beta, t) - \mathbb{E}S_n^{(m)}(\beta, t)|,$$

à cause de l'hypothèse (H2bis).

On utilise des inégalités maximales inspirées de [Pol90] plus particulièrement la proposition A.3.1 qui peut être trouvée dans l'annexe A. Avec les mêmes notations, on prend, pour t dans $[0, \tau]$ et β dans \mathcal{B}' :

$$\begin{aligned} \mathcal{F}_{n\omega} &= \{(f_{n1}(\omega, t, \beta), \dots, f_{nn}(\omega, t, \beta)), t \in [0, \tau], \beta \in \mathcal{B}'\}, \\ f_{ni}(\omega, t, \beta) &= \frac{1}{n} Y_i(t) Z_i^m(t) \exp(\beta Z_i(t)), m \in \{0, 1, 2\}. \end{aligned}$$

On a

$$\delta_{n\omega} = \sup_{f \in \mathcal{F}_{n\omega}} \|f\|_2 \leq \frac{1}{\sqrt{n}} B^m e^{B \times C(\mathcal{B}')},$$

où $C(\mathcal{B}') = \sup\{|\beta|; \beta \in \mathcal{B}'\}$.

Ici, nous reprenons essentiellement les arguments de [BGY97]. Nous considérons dans un premier temps que $m = 0$. Pour i entier positif, Y_i est une fonction croissante sur $[0, \tau]$ (donc a une pseudo-dimension 1), et par le lemme 4.1.5, \mathcal{H} est euclidienne, $\mathcal{F}_{n\omega}$ est donc euclidienne comme produit terme à terme de processus euclidiens (voir lemme A.3.3). Finalement, la somme terme à terme de processus euclidiens est euclidienne par [BGY97, Lemme A.1], ce qui nous permet de conclure.

Puisque les Z_i ont des variations totales bornées, le résultat précédent et les résultats de stabilité des lemmes A.3.2 et A.3.3 nous permettent d'obtenir le fait que $\mathcal{F}_{n\omega}$ est euclidienne pour $m = 1, 2$. Ensuite, par la proposition A.3.1 on obtient la convergence complète de $\sup_{\beta \in \mathcal{B}'} \sup_{t \in [0, \tau]} |S_n^{(m)}(\beta, t) - \mathbb{E}S_n^{(m)}(\beta, t)|$ vers 0. \square

Démonstration du lemme 4.1.2 :

On peut écrire que :

$$\begin{aligned} & \sup_{(\beta,t) \in \mathcal{B}' \times [0,\tau]} \left| \frac{S_n^{(1)}(\beta,t)}{S_n^{(0)}(\beta,t)} - \frac{s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)} \right| \\ & \leq \sup_{(\beta,t) \in \mathcal{B}' \times [0,\tau]} \left| \frac{S_n^{(1)}(\beta,t)}{S_n^{(0)}(\beta,t)} \right| \left| \frac{s^{(0)}(\beta,t) - S_n^{(0)}(\beta,t)}{s^{(0)}(\beta,t)} \right| \\ & + \sup_{(\beta,t) \in \mathcal{B}' \times [0,\tau]} \left| \frac{S_n^{(1)}(\beta,t) - s^{(1)}(\beta,t)}{s^{(0)}(\beta,t)} \right|. \end{aligned}$$

Par (H3), il existe $\eta > 0$ tel que $s^{(0)}(\beta,t) > \eta$, uniformément sur $\mathcal{B} \times [0,\tau]$. En outre, le quotient $S_n^{(1)}(\beta,t)/S_n^{(0)}(\beta,t)$ est borné par B et avec le lemme 4.1.1, on termine la preuve. \square

Démonstration du lemme 4.1.3 :

Les M_i , pour $1 \leq i \leq n$ sont des martingales, il vient donc que, pour $t \in [0,\tau]$,

$$\mathbb{E} \left(\frac{1}{n} \sum_{i=1}^n \int_0^t Z_i(s) dM_i(s) \right) = 0,$$

ce qui signifie en outre que :

$$\frac{1}{n} \sum_{i=1}^n \mathbb{E} \int_0^t Z_i(s) dN_i(s) - \int_0^t \mathbb{E} S_n^{(1)}(\beta_0, s) \lambda_0(s) ds = 0.$$

Décomposons $n^{-1} \sum_{i=1}^n \int_0^t Z_i(s) dM_i(s)$ et introduisons la somme précédente :

$$\frac{1}{n} \sum_{i=1}^n \int_0^t Z_i(s) dM_i(s) = \int_0^t (s^{(1)}(\beta_0, s) - S_n^{(1)}(\beta_0, s)) \lambda_0(s) ds \quad (4.4)$$

$$+ \int_0^t (\mathbb{E} S_n^{(1)}(\beta_0, s) - s^{(1)}(\beta_0, s)) \lambda_0(s) ds \quad (4.5)$$

$$+ \frac{1}{n} \sum_{i=1}^n \left(\int_0^t Z_i(s) dN_i(s) - \mathbb{E} \int_0^t Z_i(s) dN_i(s) \right). \quad (4.6)$$

On utilise le lemme 4.1.2 pour déduire la convergence complète de (4.4) (uniforme en la variable t) :

$$\begin{aligned} \sup_{t \in [0,\tau]} \left| \int_0^t (s^{(1)}(\beta_0, s) - S_n^{(1)}(\beta_0, s)) \lambda_0(s) ds \right| & \leq \int_0^\tau \lambda_0(s) ds \\ & \times \sup_{t \in [0,\tau]} |s^{(1)}(\beta_0, t) - S_n^{(1)}(\beta_0, t)|. \end{aligned}$$

C'est ensuite l'hypothèse (H2bis) qui nous permet de déduire la convergence complète du terme (4.5).

On applique enfin la proposition A.3.1 pour obtenir la convergence du troisième terme. À cette fin, nous devons montrer que

$$\mathcal{F}_{n\omega} = \left\{ \left(\frac{1}{n} \int_0^t Z_1(s) dN_1(s), \dots, \frac{1}{n} \int_0^t Z_n(s) dN_n(s) \right); t \in [0, \tau] \right\}$$

est euclidienne, ainsi que

$$\delta_{n\omega} = \sup_{f \in \mathcal{F}_{n\omega}} \|f\|_2 = \mathcal{O} \left(\frac{1}{\sqrt{n}} \right),$$

uniformément en ω .

Cette dernière hypothèse est directement satisfaite, d'après l'hypothèse (H4) sur les variables explicatives. De plus, les Z_i étant à variations totales bornées, on peut les écrire sous la forme $Z_i^+ - Z_i^-$, où Z_i^+ et Z_i^- sont des fonctions positives croissantes. Cela nous donne, pour $t \in [0, \tau]$,

$$\begin{aligned} \frac{1}{n} \int_0^t Z_i(s) dN_i(s) &= \frac{1}{n} Z_i^+(T_i) \delta_i \mathbb{1}_{\{X_i \leq t\}} \\ &\quad - \frac{1}{n} Z_i^-(T_i) \delta_i \mathbb{1}_{\{X_i \leq t\}}, \end{aligned}$$

qui est une somme de deux fonctions monotones, donc issues de deux familles euclidiennes (par le lemme A.3.2). Cela nous permet de déduire que la famille $\mathcal{F}_{n\omega}$ est euclidienne, comme somme terme à terme de deux familles euclidiennes (lemme A.3.1) ce qui achève la démonstration. \square

Démonstration du lemme 4.1.4 :

Il s'agit ici d'un cas plus simple que celui traité lors du lemme précédent. Pour s'y ramener sans faire de calcul, il suffit d'écrire $M_i(t) = \int_0^t 1 dM_i(s)$, où 1 joue le rôle de la variable explicative Z_i et de conclure par le lemme 4.1.3.

D'autre part, le fait que le processus soit à variations totales bornées est évident, puisque, pour $t \in [0, \tau]$,

$$\frac{1}{n} \sum_{i=1}^n M_i(t) = \frac{1}{n} \sum_{i=1}^n N_i(t) - \int_0^t \left(\frac{1}{n} \sum_{i=1}^n Y_i(s) e^{\beta Z_i(s)} \right) \lambda_0(s) ds,$$

qui est la différence de deux processus croissants positifs bornés presque sûrement. \square

Démonstration de (C1)

Procédons à présent à la démonstration de (C1). Les techniques utilisées sont essentiellement les mêmes, il s'agit de voir \mathcal{I}_n comme une intégrale de termes dont la

convergence complète peut être aisément établie. Ces termes se décomposent alors de manière simple comme produit et somme de processus, ce qui permet d'utiliser les résultats de stabilité de la convergence complète. Montrer (C1) nécessite l'acquisition préalable du lemme suivant.

Lemme 4.1.7. *Il existe un voisinage borné \mathcal{B}' de β_0 pour lequel*

$$\sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} |V_n(\beta, t) - v(\beta, t)| \xrightarrow{c} 0, \text{ lorsque } n \rightarrow \infty.$$

Montrons (C1) : à l'aide d'inégalités triangulaires, on obtient

$$\sup_{\beta \in \mathcal{B}'} |\mathcal{I}_n(\beta) - \mathbb{E}\mathcal{I}_n(\beta_0)| \leq \sup_{\beta \in \mathcal{B}'} |\mathcal{I}_n(\beta) - \mathcal{I}_n(\beta_0)| \quad (4.7)$$

$$+ |\mathcal{I}_n(\beta_0) - \mathcal{I}(\beta_0)| \quad (4.8)$$

$$+ |\mathcal{I}(\beta_0) - \mathbb{E}\mathcal{I}_n(\beta_0)|. \quad (4.9)$$

On décide de traiter les trois termes séparément. Concernant le terme (4.7), il est possible d'écrire que

$$\begin{aligned} \sup_{\beta \in \mathcal{B}'} |\mathcal{I}_n(\beta) - \mathcal{I}_n(\beta_0)| &= \sup_{\beta \in \mathcal{B}'} \left| \frac{1}{n} \sum_{i=1}^n \int_0^\tau (V_n(\beta, s) - V_n(\beta_0, s)) dN_i(s) \right| \\ &\leq \sup_{\beta \in \mathcal{B}'} \sup_{s \in [0, \tau]} |V_n(\beta, s) - v(\beta, s)| \\ &\quad + \sup_{s \in [0, \tau]} |V_n(\beta_0, s) - v(\beta_0, s)| \\ &\quad + \sup_{\beta \in \mathcal{B}'} \sup_{s \in [0, \tau]} |v(\beta_0, s) - v(\beta, s)|. \end{aligned} \quad (4.10)$$

Le lemme 4.1.7 nous fournit la convergence complète des deux premiers membres de droite de l'inégalité précédente. Par (H3), on sait que le terme déterministe (4.10) converge vers 0, lorsque β approche β_0 , il nous permet de déduire δ_ε pour lequel ce terme est inférieur à $\varepsilon/3$ et nous fournit ainsi le voisinage de β_0 à considérer. Par conséquent, on déduit la convergence complète de (4.7).

On écrit ensuite (4.8) comme suit

$$|\mathcal{I}_n(\beta_0) - \mathcal{I}(\beta_0)| \leq \left| \frac{1}{n} \int_0^\tau (V_n(\beta_0, s) - v(\beta_0, s)) d\overline{M}(s) \right| \quad (4.11)$$

$$+ \left| \frac{1}{n} \int_0^\tau v(\beta_0, s) d\overline{M}(s) \right| \quad (4.12)$$

$$+ \left| \int_0^\tau \left(V_n(\beta_0, s) \frac{S_n^{(0)}(\beta_0, s)}{n} - v(\beta_0, s) s^{(0)}(\beta_0, s) \right) \lambda_0(s) ds \right|, \quad (4.13)$$

où $\overline{M}(t) = \sum_{i=1}^n M_i(t)$.

D'après la proposition 3.1.2, on montre que le terme de droite dans (4.11) converge complètement vers 0 à l'aide des lemmes 4.1.7 et 4.1.4. Grâce aux lemmes 4.1.7 et 4.1.1 et à l'hypothèse (H1), on obtient la convergence complète de

(4.13) vers 0. Enfin, pour montrer que (4.12) converge vers 0 complètement, on utilise le corollaire B.1.1 comme cela a déjà été fait dans la démonstration de (C3). La démonstration étant exactement la même (en remplaçant $e(\beta_0, \cdot)$ par $v(\beta_0, \cdot)$), elle sera omise.

(C1) est finalement vérifiée, puisque le terme (4.9) converge lui aussi complètement, par l'hypothèse (H2bis). \square

Démonstration du lemme 4.1.7 :

Remarquons dans un premier temps que

$$\begin{aligned} & \sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} |V_n(\beta, t) - v(\beta, t)| \\ & \leq \sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} \left| \frac{S_n^{(2)}(\beta, t)}{S_n^{(0)}(\beta, t)} - \frac{s^{(2)}(\beta, t)}{s^{(0)}(\beta, t)} \right| \end{aligned} \quad (4.14)$$

$$+ \sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} \left| \left(\frac{S_n^{(1)}(\beta, t)}{S_n^{(0)}(\beta, t)} \right)^2 - \left(\frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right)^2 \right|. \quad (4.15)$$

On peut voir que (4.14) converge complètement vers 0 dans un voisinage \mathcal{B}' de β_0 . Il suffit pour cela d'adapter la preuve du lemme 4.1.2 (en notant que $S_n^{(2)}/S_n^{(0)}$ est majoré par B^2 et en remplaçant $S_n^{(1)}$ par $S_n^{(2)}$).

On souhaite ensuite écrire une majoration de (4.15) qui puisse nous permettre l'utilisation de résultats déjà obtenus, c'est pourquoi on écrit que

$$\begin{aligned} & \sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} \left| \left(\frac{S_n^{(1)}(\beta, t)}{S_n^{(0)}(\beta, t)} \right)^2 - \left(\frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right)^2 \right| \\ & \leq \sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} \left| \frac{S_n^{(1)}(\beta, t)}{S_n^{(0)}(\beta, t)} - \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right| \\ & \times \sup_{(\beta, t) \in \mathcal{B}' \times [0, \tau]} \left| \frac{S_n^{(1)}(\beta, t)}{S_n^{(0)}(\beta, t)} + \frac{s^{(1)}(\beta, t)}{s^{(0)}(\beta, t)} \right|, \end{aligned}$$

ce qui nous fournit la convergence complète de (4.15) (le second terme de l'inégalité ci-dessus étant borné), et achève la démonstration du lemme. \square

4.2 Normalité asymptotique : preuve de (C4)

ans cette partie, nous démontrons la normalité asymptotique de la fonction score, pour une taille d'échantillon T_n , où $(T_n)_{n \geq 1}$ est une suite régulière de variables aléatoires.

Rappelons que nous devons montrer que pour toute suite régulière $(T_n)_{n \geq 1}$ de variables aléatoires entières,

$$\sqrt{T_n} U_{T_n}(\beta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)).$$

Soit donc $(T_n)_{n \geq 1}$, une suite régulière de variables aléatoires, c'est-à-dire telle qu'il existe une suite déterministe $(t_n)_{n \geq 1}$ entière vérifiant :

$$\begin{cases} t_n & \rightarrow +\infty, \\ T_n/t_n & \xrightarrow{P} 1, \quad \text{lorsque } n \rightarrow \infty. \end{cases}$$

Pour montrer (C4), nous allons voir que la condition d'Anscombe (voir [Ans52] et Annexe C) est vérifiée. Il s'avère difficile de manipuler directement le terme U_{T_n} , nous définissons alors une autre suite de processus, qui lui est asymptotiquement équivalente au sens de la convergence complète et pour laquelle nous montrons la condition d'Anscombe. L'équivalence entre les deux processus nous permet enfin de conclure.

L'objet du paragraphe qui suit est de montrer que deux suites de variables aléatoires complètement asymptotiquement équivalentes satisfont la condition d'Anscombe dès lors que l'une des deux suites la satisfait.

4.2.1 L'équivalence complète et asymptotique

Définition 4.2.1. *Deux suites aléatoires $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ sont complètement asymptotiquement équivalentes si, pour tout $\varepsilon > 0$,*

$$\sum_{n \geq 1} \mathbb{P}(|X_n - Y_n| > \varepsilon) < +\infty.$$

Lemme 4.2.1. *Soit $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$, deux suites de variables aléatoires complètement asymptotiquement équivalentes. Si $(X_n)_{n \geq 1}$ converge en distribution vers X , alors $(Y_n)_{n \geq 1}$ aussi.*

Démonstration :

$(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ étant complètement asymptotiquement équivalentes, on a :

$$\forall \varepsilon > 0, \sum_{n \geq 1} \mathbb{P}(|X_n - Y_n| > \varepsilon) < \infty,$$

en particulier $\mathbb{P}(|X_n - Y_n| > \varepsilon)$ tend vers 0, et $(X_n - Y_n)_{n \geq 1}$ converge vers 0 en probabilité, donc en loi.

En décomposant Y_n , pour n entier fixé, comme somme de $Y_n - X_n$ et de X_n , en utilisant le fait que $(X_n)_{n \geq 1}$ converge en loi vers X et le théorème de Slutsky, on en déduit que $(Y_n)_{n \geq 1}$ converge en loi vers X . \square

Proposition 4.2.1. *Soit $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ deux suites complètement asymptotiquement équivalentes. Si $(X_n)_{n \geq 1}$ satisfait la condition d'Anscombe [Ans52, GMS97], et $(T_n)_{n \geq 1}$ est une suite régulière, alors,*

$$Y_{T_n} - Y_{t_n} \xrightarrow{P} 0, \text{ lorsque } n \rightarrow \infty.$$

Si, de plus, $(X_n)_{n \geq 1}$ converge vers X en distribution, alors $(Y_{T_n})_{n \geq 1}$ aussi.

Démonstration :

Soit ε un réel strictement positif. Pour n entier, on a

$$\begin{aligned} \mathbb{P}(|Y_{T_n} - Y_{t_n}| > \varepsilon) &\leq \mathbb{P}(|Y_{T_n} - X_{T_n}| > \varepsilon/3) \\ &\quad + \mathbb{P}(|X_{T_n} - X_{t_n}| > \varepsilon/3) \\ &\quad + \mathbb{P}(|X_{t_n} - Y_{t_n}| > \varepsilon/3). \end{aligned}$$

Le second terme de la majoration ci-dessus tend vers 0 par le théorème 2.7.1 de [GMS97], après des adaptations expliquées dans [GMS97, p.318].

Le troisième terme converge vers 0 car $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ sont complètement asymptotiquement équivalentes et donc pour tout ε strictement positif, $\mathbb{P}(|X_n - Y_n| > \varepsilon)$ est le terme général d'une série convergente.

Si δ est un réel strictement positif de $]0, 1[$,

$$\begin{aligned} \mathbb{P}(|Y_{T_n} - X_{T_n}| > \varepsilon/3) &\leq \mathbb{P}(|Y_{T_n} - X_{T_n}| > \varepsilon/3; T_n \geq t_n(1 - \delta)) \\ &\quad + \mathbb{P}(T_n < t_n(1 - \delta)) \\ &\leq \sum_{k \geq (1-\delta)t_n} \mathbb{P}(|Y_k - X_k| > \varepsilon/3) + \mathbb{P}(T_n/t_n < 1 - \delta). \end{aligned}$$

La première quantité n'est autre que le reste d'une série convergente, et la seconde tend vers 0 grâce à la régularité de la suite $(T_n)_{n \geq 1}$. On a donc la convergence en probabilité de $(Y_{T_n} - Y_{t_n})_{n \geq 1}$ vers 0.

Pour le dernier point, il nous suffit d'écrire $Y_{T_n} = (Y_{T_n} - Y_{t_n}) + Y_{t_n}$, pour n entier fixé, puis d'utiliser le théorème de Slutsky et le fait que $Y_{t_n} \xrightarrow{D} X$ (les suites $(Y_{t_n})_{n \geq 1}$ et $(X_{t_n})_{n \geq 1}$ étant complètement asymptotiquement équivalentes, on applique le lemme 4.2.1) et que $Y_{T_n} - Y_{t_n} \xrightarrow{P} 0$. \square

4.2.2 Démonstration de (C4)

On définit \tilde{U}_n , pour $t \in [0, \tau]$, par :

$$\tilde{U}_n(\beta_0, t) = \frac{1}{n} \sum_{i=1}^n \int_0^t Z_i(s) dM_i(s) - \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{s^{(1)}(s)}{s^{(0)}(s)} dM_i(s).$$

Nous allons voir qu'il est plus aisé de montrer la condition d'Anscombe et de ce fait la normalité asymptotique pour la suite $(\sqrt{n}\tilde{U}_n)_{n \geq 1}$, puisqu'elle se présente sous la forme de sommes de processus indépendants.

Lemme 4.2.2. *Les suites $(\sqrt{n}U_n)_{n \geq 1}$ et $(\sqrt{n}\tilde{U}_n)_{n \geq 1}$ sont asymptotiquement équivalentes en probabilité.*

Démonstration :

Pour $t \in [0, \tau]$, on écrit

$$\begin{aligned} \sqrt{n} \left(U_n(\beta_0, t) - \tilde{U}_n(\beta_0, t) \right) &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t \left(\frac{S_n^{(1)}(\beta_0, s)}{S_n^{(0)}(\beta_0, s)} - \frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} \right) dM_i(s) \\ &= -\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t (E_n(s) - e(s)) dM_i(s), \end{aligned}$$

qui est une martingale dont le processus des variations prévisibles est donné par

$$\begin{aligned} n \left\langle U_n(\beta_0, \cdot) - \tilde{U}_n(\beta_0, \cdot) \right\rangle (t) &= \frac{1}{n} \sum_{i=1}^n \int_0^t (E_n(s) - e(s))^2 d \langle M_i \rangle (s) \\ &= \int_0^t (E_n(s) - e(s))^2 S_n^{(0)}(s) \lambda_0(s) ds. \end{aligned}$$

Par ailleurs :

$$\begin{aligned} n \text{Var}(U_n(\beta_0, t) - \tilde{U}_n(\beta_0, t)) &= n \mathbb{E} \left\langle U_n(\beta_0, \cdot) - \tilde{U}_n(\beta_0, \cdot) \right\rangle (t) \\ &\leq A \mathbb{E} \left(\sup_{t \in [0, \tau]} |E_n(t) - e(t)| \right), \end{aligned}$$

où A est une constante positive.

On en déduit que la variance précédente tend vers 0 (voir le lemme 4.1.2). Il nous suffit ensuite de voir que, comme différence de martingales, l'espérance du processus $\sqrt{n} \left(U_n(\beta_0, \cdot) - \tilde{U}_n(\beta_0, \cdot) \right)$ est nulle, pour déduire qu'il tend vers 0, en probabilité. \square

Ce résultat est toutefois insuffisant pour pouvoir appliquer la proposition 4.2.1, mais il suggère de démontrer le résultat suivant, qui est plus fort.

Lemme 4.2.3. *Les suites $(\sqrt{n}U_n)_{n \geq 1}$ et $(\sqrt{n}\tilde{U}_n)_{n \geq 1}$ sont complètement asymptotiquement équivalentes.*

Démonstration :

Soit ε , un réel strictement positif. Montrons que la série de terme général $\mathbb{P}(|\sqrt{n}U_n(\beta_0, t) - \sqrt{n}\tilde{U}_n(\beta_0, t)| > \varepsilon)$ est convergente, pour β_0 et t fixés.

Pour ce faire, nous utilisons une inégalité maximale pour martingales proposée dans [SW86, p.900], qui fait l'objet du théorème B.1.1 de l'annexe B et nous appliquons le corollaire B.1.1 en prenant $M_n = \sqrt{n} (U_n - \tilde{U}_n)$.

On vérifie d'abord que M est une martingale locale de carré intégrable, qu'elle est uniformément intégrable, et que l'on peut majorer son processus de saut uniformément en t , par c/\sqrt{n} , où c est une constante positive (voir démonstration de (C3), page 66).

La première condition du corollaire B.1.1 est donc clairement vérifiée. Pour la seconde condition du corollaire, on utilise l'inégalité maximale de la proposition A.3.1 (on a déjà vu que les conditions pour obtenir l'inégalité maximale étaient vérifiées), et on obtient (B, c_1 et c_2 étant des constantes positives) :

$$\begin{aligned} & \mathbb{P} \left(\left\langle \sqrt{n} (U_n - \tilde{U}_n) \right\rangle (t) \geq n^{-1/4} \right) \\ &= \mathbb{P} \left(\int_0^t (E_n(s) - e(s))^2 S^{(0)}(s) \lambda_0(s) ds \geq n^{-1/4} \right) \\ &\leq \mathbb{P} \left(\sup_{t \in [0, \tau]} |E_n(t) - e(t)| \geq B n^{-1/8} \right) \\ &\leq c_1 \exp(-c_2 n^{3/4}). \end{aligned}$$

On a encore une fois le terme général d'une série convergente. On obtient alors le résultat souhaité. \square

Lemme 4.2.4. *La suite $(\sqrt{n}\tilde{U}_n)_{n \geq 1}$ satisfait la condition d'Anscombe, i.e.*

$$\forall \varepsilon > 0, \exists \delta > 0; \mathbb{P} \left(\max_{0 \leq k \leq \delta n} |\sqrt{n+k}\tilde{U}_{n+k} - \sqrt{n}\tilde{U}_n| > \varepsilon \right) < \varepsilon.$$

De plus, on a :

$$\sqrt{n}\tilde{U}_n(\beta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)).$$

Démonstration :

Soit $\varepsilon > 0$ et k un entier.

Pour tout t dans $[0, \tau]$ (en omettant une nouvelle fois β_0), on a

$$\begin{aligned}
\left| \sqrt{n+k} \tilde{U}_{n+k}(t) - \sqrt{n} \tilde{U}_n(t) \right| &= \left| \frac{1}{\sqrt{n+k}} \sum_{i=1}^{n+k} \int_0^t (Z_i(s) - e(s)) dM_i(s) \right. \\
&\quad \left. - \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| \\
&\leq \left| \left(\frac{1}{\sqrt{n+k}} - \frac{1}{\sqrt{n}} \right) \sum_{i=1}^n \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| \\
&\quad + \left| \frac{1}{\sqrt{n+k}} \sum_{i=n+1}^{n+k} \int_0^t (Z_i(s) - e(s)) dM_i(s) \right|.
\end{aligned}$$

On traite les deux termes ci-dessus séparément.

Premier terme : définissons l'expression suivante, pour n , t et ε fixés :

$$P_{n,t,\varepsilon}^{(1)} = \mathbb{P} \left(\max_{0 \leq k \leq n\delta} \left| \left(\frac{1}{\sqrt{n+k}} - \frac{1}{\sqrt{n}} \right) \sum_{i=1}^n \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| > \frac{\varepsilon}{2} \right).$$

Isolons dans un premier temps le terme à l'intérieur du maximum afin de le rendre indépendant de k :

$$\begin{aligned}
&\max_{0 \leq k \leq n\delta} \left| \left(\frac{1}{\sqrt{n+k}} - \frac{1}{\sqrt{n}} \right) \sum_{i=1}^n \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| \\
&\leq \max_{0 \leq k \leq n\delta} \frac{1}{\sqrt{n}} \left(1 - \frac{1}{\sqrt{\delta+1}} \right) \left| \sum_{i=1}^n \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| \\
&\leq \frac{1}{\sqrt{n}} \left(1 - \frac{1}{\sqrt{\delta+1}} \right) \left| \sum_{i=1}^n \int_0^t (Z_i(s) - e(s)) dM_i(s) \right|,
\end{aligned}$$

d'où

$$P_{n,t,\varepsilon}^{(1)} \leq \mathbb{P} \left(\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| > C_{\delta,\varepsilon} \right),$$

avec $C_{\delta,\varepsilon} = \frac{\varepsilon}{2} \left(1 - \frac{1}{\sqrt{\delta+1}} \right)^{-1}$.

On applique le théorème de Rebolledo (voir par exemple [AG82, Appendix I p.1115, théorème I.2]). On sait que l'intensité de $N_i(\cdot)$ est donnée par

$$\lambda_i(t) = \lambda_0(t) Y_i(t) \exp(\beta_0 Z_i(t)).$$

Posons

$$H_i(t) = \frac{1}{\sqrt{n}} (Z_i(t) - e(t)),$$

alors

$$\sqrt{n}\tilde{U}_n(t) = \int_0^t \sum_{l=1}^n H_l(u) dM_l(u).$$

Vérifions les deux hypothèses nécessaires à l'application du théorème de Rebolledo.

Hypothèse 1 : $\langle \sqrt{n}\tilde{U}_n, \sqrt{n}\tilde{U}_n \rangle (t) \xrightarrow{P} A(t)$, lorsque $n \rightarrow +\infty$, pour $t \in [0, \tau]$.

Pour t élément de $[0, \tau]$, on peut écrire :

$$\begin{aligned} n \langle \tilde{U}_n, \tilde{U}_n \rangle (t) &= \int_0^t \sum_{l=1}^n H_l(u)^2 \lambda_l(u) du \\ &= \frac{1}{n} \int_0^t \sum_{l=1}^n (Z_l(u)^2 + e(u)^2 - Z_l(u)e(u)) \lambda_0(u) Y_l(u) e^{(\beta_0 Z_l(u))} du \\ &= \int_0^t (S_n^{(2)}(u) + S_n^{(0)}(u)e(u)^2 - 2S_n^{(1)}(u)e(u)) \lambda_0(u) du \\ &\xrightarrow{P} \int_0^t (s^{(2)}(u) - s^{(1)}(u)e(u)) \lambda_0(u) du. \end{aligned}$$

En effet d'après les lemmes 4.1.1 et 4.1.2, $(S_n^{(2)} + S_n^{(0)}e^2 - 2S_n^{(1)}e)_{n \geq 1}$ converge complètement vers $(s^{(2)} - s^{(1)}e)$ sur l'intervalle d'intégration qui est borné.

Hypothèse 2 : $\forall \varepsilon > 0, \int_0^\tau \sum_{l=1}^n H_l(t)^2 \lambda_l(t) \mathbf{1}_{\{|H_l(t)| > \varepsilon\}} dt \xrightarrow{P} 0$, lorsque $n \rightarrow +\infty$.

Soit ε , un réel strictement positif.

$$\int_0^\tau \sum_{l=1}^n H_l(t)^2 \lambda_l(t) \mathbf{1}_{\{|H_l(t)| > \varepsilon\}} dt = \frac{1}{n} \int_0^\tau \sum_{l=1}^n (Z_l(t) - e(t))^2 \lambda_l(t) \mathbf{1}_{\{\frac{1}{\sqrt{n}}|Z_l(t) - e(t)| > \varepsilon\}} dt.$$

Intéressons-nous à l'indicatrice : on sait que les variables explicatives Z_l sont bornées indépendamment de l et que e est bornée, on en déduit que l'indicatrice est nulle à partir d'un certain rang, ce qui démontre le fait que l'intégrale tend vers 0, en probabilité.

Par le théorème de Rebolledo, on obtient la deuxième partie du lemme, *i.e.*

$$\sqrt{n}\tilde{U}_n(\beta_0) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)).$$

De plus, on sait que $(\sqrt{n}\tilde{U}_n)_{n \geq 1}$ converge vers une loi gaussienne, dont la fonction de répartition est notée Φ . On a

$$P_{n,t,\varepsilon}^{(1)} \leq \mathbb{P}(|\sqrt{n}\tilde{U}_n(t)| > C_{\delta,\varepsilon}) \xrightarrow[n \rightarrow \infty]{} 2(1 - \Phi(C_{\delta,\varepsilon})),$$

avec $C_{\delta,\varepsilon} \xrightarrow{\delta \rightarrow 0^+} +\infty$. Soit $\delta = \delta_1 > 0$ tel que $2(1 - \Phi(C_{\delta,\varepsilon})) < \varepsilon/2$, on obtient alors $P_{n,t,\varepsilon}^{(1)} < \varepsilon/2$.

Deuxième terme : nous adoptons une notation semblable pour ce terme en définissant

$$P_{n,t,\varepsilon}^{(2)} = \mathbb{P} \left(\max_{0 \leq k \leq n\delta} \left| \frac{1}{\sqrt{n+k}} \sum_{i=n+1}^{n+k} \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| > \frac{\varepsilon}{2} \right).$$

On utilise l'inégalité de Kolmogorov [HJ94, p.472] et l'on obtient :

$$\begin{aligned} P_{n,t,\varepsilon}^{(2)} &\leq \mathbb{P} \left(\max_{0 \leq k \leq n\delta} \left| \sum_{i=n+1}^{n+k} \int_0^t (Z_i(s) - e(s)) dM_i(s) \right| > \frac{\varepsilon\sqrt{n}}{2} \right) \\ &\leq \frac{4}{\varepsilon^2 n} \sum_{i=n+1}^{n+[n\delta]} \text{Var} \left(\int_0^t (Z_i(s) - e(s)) dM_i(s) \right) \\ &\leq \frac{4C}{\varepsilon^2 n} \sum_{i=n+1}^{n+[n\delta]} \mathbb{E} \left(\int_0^t d\langle M_i \rangle(s) \right) \\ &\leq \frac{B\delta}{\varepsilon^2}, \end{aligned}$$

où B et C sont des constantes.

On choisit $\delta = \min\{\delta_1, \delta_2\}$, avec δ_2 , un réel strictement positif tel que $\delta_2 < \varepsilon^3/B$. On obtient alors la condition d'Anscombe désirée. \square

Maintenant, en appliquant le lemme 4.2.1, on sait que les suites $(\sqrt{n}U_n)_{n \geq 1}$ et $(\sqrt{n}\tilde{U}_n)_{n \geq 1}$ sont complètement asymptotiquement équivalentes (par le lemme 4.2.3). De plus, $(\sqrt{n}\tilde{U}_n)_{n \geq 1}$ satisfait la condition d'Anscombe (par le lemme 4.2.4), on conclut par la proposition 4.2.1.

Chapitre 5

Normalité asymptotique de la variable d'arrêt N_d

Sommaire

5.1	Résultats préliminaires	83
5.2	Hypothèses et représentations de $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$	86
5.2.1	Notations	86
5.2.2	Hypothèses	87
5.3	Normalité asymptotique de $\sqrt{N_d} - \sqrt{n_d}$	91
5.3.1	Démonstration du théorème 5.3.1	91

Par un théorème issu de [GMS97], nous verrons que montrer la normalité asymptotique de la variable d'arrêt se ramène à montrer celle de l'estimateur de la variance asymptotique du paramètre de régression β_0 , pour une taille d'échantillon donnée par une suite régulière de règle d'arrêt. Ainsi, deux conditions nous permettront d'obtenir cette propriété : la condition d'Anscombe pour l'estimateur de la variance asymptotique et sa normalité asymptotique.

Nous verrons dans un premier temps quelques résultats préliminaires permettant d'établir la condition d'Anscombe, puis nous définirons de nouvelles notations permettant de donner des hypothèses sur de nouveaux processus qui s'ajouteront à (H1-H4). En outre, nous nous placerons au cours de cette partie dans le cas d'observations indépendantes et identiquement distribuées (ces hypothèses seront définies dans la seconde partie de ce chapitre). Par la suite, nous donnons des représentations équivalentes au sens de la convergence complète puis de la convergence en loi de la variable aléatoire $\sqrt{n} \left(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau) \right)$.

Enfin, au cours de la troisième partie, nous démontrons les conditions suffisantes à la convergence en loi de la variable d'arrêt.

5.1 Résultats préliminaires

On souhaite établir un résultat de convergence en loi pour $\sqrt{N_d} - \sqrt{n_d}$. À cette fin, nous rappelons dans un premier temps un résultat obtenu dans [GMS97] dans le cas d'une variable d'arrêt N_d définie pour $d > 0$ par

$$N_d = \min \left\{ n \geq n_0; n \geq \frac{u^2 \hat{\sigma}_n^2}{d^2} \right\}. \quad (5.1)$$

On définit également la taille idéale de l'échantillon n_d par :

$$n_d = \min \left\{ n \geq n_0; n \geq \frac{u^2 \sigma^2}{d^2} \right\}.$$

On dispose alors de la proposition suivante.

Proposition 5.1.1. *Supposons que*

1. $\hat{\sigma}_n^2 \xrightarrow{p.s.} \sigma^2 > 0$, lorsque $n \rightarrow +\infty$,
2. Il existe un nombre réel $\gamma^2 > 0$ tel que : $\sqrt{n_d}(\hat{\sigma}_{N_d}^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2)$, lorsque $d \rightarrow 0^+$, pour N_d définie par (5.1).

Alors, on a :

$$\sqrt{N_d} - \sqrt{n_d} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\gamma^2}{4\sigma^2}\right),$$

lorsque $d \rightarrow 0^+$.

Démonstration : voir [GMS97, p.321]. □

Pour obtenir un résultat similaire, il nous suffit de montrer l'hypothèse 2. En effet, l'hypothèse 1. a déjà été établie au cours du corollaire 3.2.1 p. 60. Dans un premier temps, nous donnons quelques résultats préliminaires qui concernent la condition d'Anscombe.

Lemme 5.1.1. *Soit $(X_i)_{i \geq 1}$ une suite de variables aléatoires indépendantes, identiquement distribuées et de variance finie σ^2 . Alors*

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}X_i)$$

satisfait la condition d'Anscombe.

Démonstration : Pour plus de clarté, considérons des variables X_i centrées. Nous procédons ici de la même façon que celle utilisée lors de la démonstration du lem-

me 4.2.4. Il est facile de montrer que, pour $\varepsilon > 0$:

$$\begin{aligned}
& \mathbb{P} \left(\max_{0 \leq k \leq \delta n} \left| \frac{1}{\sqrt{n+k}} \sum_{i=1}^{n+k} X_i - \frac{1}{\sqrt{n}} \sum_{i=1}^n X_i \right| > \varepsilon \right) \\
& \leq \mathbb{P} \left(\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n X_i \right| \times \max_{0 \leq k \leq \delta n} \left| \sqrt{\frac{n}{n+k}} - 1 \right| > \varepsilon/2 \right) \\
& + \mathbb{P} \left(\max_{0 \leq k \leq \delta n} \frac{1}{\sqrt{n+k}} \left| \sum_{i=n}^{n+k} X_i \right| > \varepsilon/2 \right) \\
& \leq \mathbb{P} \left(\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n X_i \right| > \frac{\varepsilon}{2 \left(1 - \sqrt{\frac{1}{1+\delta}}\right)} \right) + \mathbb{P} \left(\max_{0 \leq k \leq \delta n} \left| \sum_{i=n}^{n+k} X_i \right| > \varepsilon \sqrt{n}/2 \right).
\end{aligned}$$

Or, par l'inégalité de Kolmogorov (on sait que les X_i sont centrées et indépendantes), on a :

$$\mathbb{P} \left(\max_{0 \leq k \leq \delta n} \left| \sum_{i=n}^{n+k} X_i \right| > \varepsilon \sqrt{n}/2 \right) \leq \frac{\delta n \sigma^2}{(\varepsilon \sqrt{n}/2)^2} = \frac{4 \delta \sigma^2}{\varepsilon^2}.$$

Soit $\delta_1 > 0$ tel que $4\delta_1 \sigma^2 / \varepsilon^2 < \varepsilon/2$.

D'après le théorème de la limite centrale, $\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n X_i \right|$ converge en loi vers $|X|$ où X admet la loi $\mathcal{N}(0, \sigma^2)$. On note

$$C(\varepsilon, \delta) = \frac{\varepsilon}{2 \left(1 - \sqrt{\frac{1}{1+\delta}}\right)} \rightarrow +\infty, \text{ si } \delta \rightarrow 0^+.$$

Soit δ_2 tel que $2\Phi(C(\varepsilon, \delta_2)) - 1 < \varepsilon/4$ (Φ est la fonction de répartition d'une loi normale centrée réduite), il existe $n_0 \geq 1$ tel que pour tout $n \geq n_0$ on ait :

$$\mathbb{P} \left(\frac{1}{\sqrt{n}} \left| \sum_{i=1}^n X_i \right| > C(\varepsilon, \delta_2) \right) < \varepsilon/2,$$

alors pour $n \geq n_0$ et $\delta = \min(\delta_1, \delta_2)$, on obtient le résultat voulu. \square

Lemme 5.1.2. Soient deux suites de variables aléatoires $(X_n)_{n \geq 1}$ et $(Y_n)_{n \geq 1}$ telles que

1. $(X_n)_{n \geq 1}$ converge complètement vers x ;
2. $(Y_n)_{n \geq 1}$ converge en loi vers Y dont la mesure associée est diffuse ;
3. $(Y_n)_{n \geq 1}$ vérifie la condition d'Anscombe ;

alors, $(Y_n X_n)_{n \geq 1}$ vérifie la condition d'Anscombe.

Démonstration : Sans perte de généralité, on peut supposer $x > 0$. Pour ε et δ deux réels strictement positifs, $n \geq 1$, et k entier compris entre 0 et δn , on écrit :

$$\begin{aligned} |X_{n+k} - X_n| &\leq |X_{n+k}| + |X_n| \\ &\leq \max_{0 \leq k \leq \delta n} |X_{n+k}| + |X_n|, \end{aligned}$$

donc on a :

$$\mathbb{P} \left(\max_{0 \leq k \leq \delta n} |X_{n+k} - X_n| > \varepsilon \right) \leq 2 \sum_{k \geq 0} \mathbb{P} (|X_{n+k}| > \varepsilon),$$

qui est le reste d'une série convergente, donc qui tend vers 0. Cela nous permet de déduire que $(X_n)_{n \geq 1}$ satisfait la condition d'Anscombe, il en va donc de même pour $(X_n - x)_{n \geq 1}$.

Remarquons les inclusions suivantes, pour $a > 0$:

$$\begin{aligned} \{|X_{n+k}Y_{n+k} - X_nY_n| > \varepsilon\} &\subset \{|Y_n(X_{n+k} - X_n)| > \varepsilon/2\} \\ &\cup \{|X_{n+k}(Y_{n+k} - Y_n)| > \varepsilon/2\} \\ &\subset \{|Y_n| > a, |Y_n(X_{n+k} - X_n)| > \varepsilon/2\} \\ &\cup \{|Y_n| \leq a, |Y_n(X_{n+k} - X_n)| > \varepsilon/2\} \\ &\cup \{|X_{n+k}| > 2x, |X_{n+k}(Y_{n+k} - Y_n)| > \varepsilon/2\} \\ &\cup \{|X_{n+k}| \leq 2x, |X_{n+k}(Y_{n+k} - Y_n)| > \varepsilon/2\} \\ &\subset \{|Y_n| > a\} \cup \left\{ \max_{0 \leq k \leq \delta n} |X_{n+k} - X_n| > \varepsilon/(2a) \right\} \\ &\cup \left\{ \max_{0 \leq k \leq \delta n} |X_{n+k}| > 2x \right\} \\ &\cup \left\{ \max_{0 \leq k \leq \delta n} |Y_{n+k} - Y_n| > \varepsilon/(4x) \right\} \end{aligned}$$

On sait que $(Y_n)_{n \geq 1}$ converge en loi vers Y dont la loi associée est diffuse, cela nous permet ainsi de choisir un réel strictement positif a (qui sera désormais fixé) et un rang n_0 à partir duquel on a :

$$\mathbb{P} (|Y_n| > a) \leq \varepsilon/4.$$

Par ailleurs, on a montré que $(X_n)_{n \geq 1}$ satisfait la condition d'Anscombe, on en déduit ainsi l'existence de n_1 et δ_1 tel que pour $n \geq n_1$, on ait (on sait que $\varepsilon/2a > 0$) :

$$\mathbb{P} \left(\max_{0 \leq k \leq \delta_1 n} |X_{n+k} - X_n| > \varepsilon/(2a) \right) < \varepsilon/4.$$

Ensuite, on remarque que :

$$\mathbb{P} \left(\max_{0 \leq k \leq \delta n} |X_{n+k}| > 2x \right) \leq \sum_{k \geq 0} \mathbb{P} (|X_{n+k}| > 2x),$$

qui est le reste d'une série convergente, par l'hypothèse 1.; il est donc possible de trouver un rang n_3 , à partir duquel cette quantité est inférieure à $\varepsilon/4$.

Enfin, par l'hypothèse 3., on sait que $(Y_n)_{n \geq 1}$ satisfait la condition d'Anscombe. On en déduit l'existence de n_2 et δ_2 tels que pour $n \geq n_2$, on ait :

$$\mathbb{P} \left(\max_{0 \leq k \leq \delta_2 n} |Y_{n+k} - Y_n| > \varepsilon/(4x) \right) < \varepsilon/4.$$

On conclut que

$$\mathbb{P} \left(\max_{0 \leq k \leq \delta n} |X_{n+k} Y_{n+k} - X_n Y_n| > \varepsilon \right) < \varepsilon,$$

pour $\delta = \min(\delta_1, \delta_2)$ et $n \geq \max(n_1, n_2, n_3)$. □

5.2 Hypothèses et représentations de $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$

Le calcul de la loi asymptotique de $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$ nécessite de faire quelques hypothèses supplémentaires sur le modèle, que nous donnons au cours de cette partie.

5.2.1 Notations

Nous introduisons dans un premier temps les nouveaux objets qui interviendront dans les hypothèses supplémentaires. Dans toute la suite, nous nous plaçons dans le cas plus restreint où les triplets $(X_i, \delta_i, Z_i(\cdot))$ sont indépendants et identiquement distribués. Dans ce cas, on a, pour $m = 0, 1, 2$, β et t donnés,

$$s^{(m)}(\beta, t) = \mathbb{E} S^{(m)}(\beta, t).$$

On définit de plus deux nouveaux processus notés $S_n^{(3)}$ et W_n pour $t \in [0, \tau]$ et $\beta \in \mathbb{R}$ par :

$$\begin{aligned} S_n^{(3)}(\beta, t) &= \frac{1}{n} \sum_{i=1}^n Y_i(t) Z_i(t)^3 \exp(\beta Z_i(t)); \\ W_n(\beta, t) &= \frac{\partial V_n}{\partial \beta}(\beta, t). \end{aligned}$$

On peut constater que l'on a la formulation suivante :

$$W_n(\beta, t) = \frac{S_n^{(3)}(\beta, t)}{S_n^{(0)}(\beta, t)} - 3 \frac{S_n^{(2)}(\beta, t) S_n^{(1)}(\beta, t)}{(S_n^{(0)}(\beta, t))^2} + 2 \left(\frac{S_n^{(1)}(\beta, t)}{S_n^{(0)}(\beta, t)} \right)^3.$$

Afin de définir la variance asymptotique, nous adoptons les notations suivantes, pour $1 \leq i \leq n$,

$$H_i(\beta_0, \tau) = \int_0^\tau m_i(\beta_0, s) dM_i(s) + \sum_{k=0}^2 \int_0^\tau \theta_k(\beta_0, s) (Z_i(t)^k Y_i(s) e^{\beta_0 Z_i} - s^{(k)}(\beta_0, s)) \lambda_0(s) ds,$$

avec

$$\begin{aligned} m_i(\beta_0, s) &= (Z_i(s) - e(\beta_0, s)) g(\beta_0, \tau) + v(\beta_0, s), \\ g(\beta_0, \tau) &= \mathcal{I}(\beta_0, \tau)^{-1} \int_0^\tau w(\beta_0, s) s^{(0)}(\beta_0, s) \lambda_0(s) ds, \\ w(\beta_0, t) &= \frac{s^{(3)}(\beta_0, t)}{s^{(0)}(\beta_0, t)} - 3 \frac{s^{(2)}(\beta_0, t) s^{(1)}(\beta_0, t)}{(s^{(0)}(\beta_0, t))^2} + 2 \left(\frac{s^{(1)}(\beta_0, t)}{s^{(0)}(\beta_0, t)} \right)^3, \\ \theta_0(\beta_0, s) &= v(\beta_0, s) - \frac{s^{(2)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} + 2 \left(\frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} \right)^2, \\ \theta_1(\beta_0, s) &= -2 \frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)}, \\ \theta_2(\beta_0, s) &= 1. \end{aligned}$$

5.2.2 Hypothèses

Hypothèses 3. *Considérons les hypothèses suivantes sur le modèle. Aux hypothèses (H1) et (H4) s'ajoutent :*

H3bis - *En conservant les notations précédentes, pour β dans \mathcal{B} , $t \in [0, \tau]$, on a :*

$$s^{(3)}(\beta, t) = \frac{\partial^3}{\partial \beta^3} s^{(0)}(\beta, t),$$

$s^{(3)}(\cdot, t)$ est une fonction continue sur \mathcal{B} , uniformément en $t \in [0, \tau]$.

H5 - *Les triplets $(X_i, \delta_i, Z_i(\cdot))$ sont indépendants et identiquement distribués.*

H6 - *Pour $1 \leq i \leq n$, on suppose que $\mathbb{E}(H_i(\beta_0, \tau))^2 = \gamma^2(\beta_0, \tau)$ est strictement positif.*

Il est possible de se ramener à une expression plus simple grâce au lemme qui suit. Nous utilisons ici l'équivalence au sens de la convergence complète, qui n'est pas nécessaire pour obtenir une formulation permettant le calcul de la loi asymptotique. Cependant, elle sera utile pour la suite, puisqu'elle permet de se ramener à une expression plus simple pour montrer la condition d'Anscombe, où l'équivalence complète se révèle cette fois nécessaire.

Lemme 5.2.1. *Sous les hypothèses (H1), (H3bis) et (H4-H6), $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$ est complètement asymptotiquement équivalent à :*

$$\varepsilon_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau V_n(\hat{\beta}_n, s) dM_i(s) \quad (5.2)$$

$$+ (\mathcal{I}(\beta_0, \tau))^{-1} \sum_{i=1}^n \int_0^\tau W_n(\beta_n^*, s) S_n^{(0)}(\beta_0, s) \lambda_0(s) ds \sqrt{n} U_n(\beta_0, \tau) \quad (5.3)$$

$$+ \sqrt{n} \sum_{k=0}^2 \int_0^\tau \theta_k(\beta_0, s) (S_n^{(k)}(\beta_0, s) - s^{(k)}(\beta_0, s)) \lambda_0(s) ds,$$

où β_n^* appartient au segment d'extrémités β_0 et $\hat{\beta}_n$.

Démonstration : Nous avons :

$$\begin{aligned} & \sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau)) \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau V_n(\hat{\beta}_n, s) dM_i(s) \\ &+ \sqrt{n} \int_0^\tau \left(V_n(\hat{\beta}_n, s) S_n^{(0)}(\beta_0, s) - v(\beta_0, s) s^{(0)}(\beta_0, s) \right) \lambda_0(s) ds \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau V_n(\hat{\beta}_n, s) dM_i(s) \\ &+ \sqrt{n} \int_0^\tau \left(V_n(\beta_0, s) S_n^{(0)}(\beta_0, s) - v(\beta_0, s) s^{(0)}(\beta_0, s) \right) \lambda_0(s) ds \\ &+ \sqrt{n}(\hat{\beta}_n - \beta_0) \int_0^\tau W_n(\beta_n^*, s) S_n^{(0)}(\beta_0, s) \lambda_0(s) ds \\ &= \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau V_n(\hat{\beta}_n, s) dM_i(s) \\ &+ (\mathcal{I}(\beta_0, \tau))^{-1} \int_0^\tau W_n(\beta_n^*, s) S_n^{(0)}(\beta_0, s) \lambda_0(s) ds \sqrt{n} U_n(\beta_0, \tau) \\ &+ \sqrt{n} \sum_{k=0}^2 \int_0^\tau \Theta_n^{(k)}(\beta_0, s) (S_n^{(k)}(\beta_0, s) - s^{(k)}(\beta_0, s)) \lambda_0(s) ds, \quad (5.4) \end{aligned}$$

où

$$\Theta_n^{(0)}(\beta_0, s) = V_n(\beta_0, s) - \frac{s^{(2)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} + \left(\frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} \right)^2 + \frac{s^{(1)}(\beta_0, s) S_n^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s) S_n^{(0)}(\beta_0, s)},$$

$$\Theta_n^{(1)}(\beta_0, s) = -\frac{S_n^{(1)}(\beta_0, s)}{S_n^{(0)}(\beta_0, s)} - \frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)},$$

$$\Theta_n^{(2)}(\beta_0, s) = 1.$$

Il nous reste maintenant à démontrer qu'il est possible de remplacer les $\Theta_n^{(k)}$ par les $\theta_n^{(k)}$ dans (5.4) au sens de l'équivalence complète. Pour cela, remarquons que l'on conserve la convergence complète de $\Delta_m = \sup_{s \in [0, \tau]} |S_n^{(m)}(\beta_0, t) - s^{(m)}(\beta_0, t)|$, $m = 0, 1, 2$ si l'on remplace Δ_m par $n^{1/4} \Delta_m$ (voir la démonstration de la proposition A.3.1).

Par conséquent, pour terminer la démonstration, il est suffisant de démontrer que pour $k = 0$ et 1 , on a :

$$n^{1/4} \sup_{s \in [0, \tau]} |\Theta_n^{(k)}(\beta_0, s) - \theta^{(k)}(\beta_0, s)| \xrightarrow{c} 0,$$

ce qui au vu des termes $\Theta_n^{(k)}$ est équivalent à montrer que :

$$n^{1/4} \sup_{s \in [0, \tau]} \left| \frac{S_n^{(1)}(\beta_0, s)}{S_n^{(0)}(\beta_0, s)} - \frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} \right| \xrightarrow{c} 0.$$

Or, on a :

$$\begin{aligned} \frac{S_n^{(1)}(\beta_0, s)}{S_n^{(0)}(\beta_0, s)} - \frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} &= \frac{S_n^{(1)}(\beta_0, s) - s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} \\ &\quad - \frac{S_n^{(1)}(\beta_0, s)}{S_n^{(0)}(\beta_0, s) s^{(0)}(\beta_0, s)} (S_n^{(0)}(\beta_0, s) - s^{(0)}(\beta_0, s)), \end{aligned}$$

et l'on sait que $\inf_{s \in [0, \tau]} s^{(0)}(\beta_0, s) = \alpha > 0$, ainsi que $\frac{S_n^{(1)}(\beta_0, s)}{S_n^{(0)}(\beta_0, s)} \leq B$ sur $[0, \tau]$, où B est une constante positive. Il s'ensuit

$$n^{1/4} \sup_{s \in [0, \tau]} \left| \frac{S_n^{(1)}(\beta_0, s)}{S_n^{(0)}(\beta_0, s)} - \frac{s^{(1)}(\beta_0, s)}{s^{(0)}(\beta_0, s)} \right| \leq \frac{1}{\alpha} n^{1/4} \Delta_1 + \frac{B}{\alpha} n^{1/4} \Delta_0,$$

qui converge complètement vers 0. □

Le lemme suivant apporte une expression équivalente au sens de la convergence en loi qui nous permettra de simplifier la démonstration de la normalité asymptotique du terme $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$.

Lemme 5.2.2. *Sous les hypothèses (H1), (H3bis) et (H4-H6), $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$ a même comportement asymptotique que*

$$\begin{aligned} &\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau m_i(\beta_0, s) dM_i(s) \\ &+ \sum_{k=0}^2 \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \theta_k(\beta_0, s) ((Z_i(t))^k Y_i(s) e^{\beta_0 Z_i} - s^{(k)}(\beta_0, s)) ds. \end{aligned}$$

Démonstration : Pour montrer ce résultat, nous appliquons le lemme 5.2.1 qui nous permet de nous ramener à une expression plus simple. Notre but est d'écrire un terme équivalent à $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0; \tau))$ au sens de la convergence en probabilité comme somme de variables aléatoires indépendantes et identiquement distribuées. Nous pourrions ensuite déduire la normalité asymptotique comme une conséquence du théorème de la limite centrale.

Dans un premier temps, on remarque que le terme (5.3) se ramène à une somme de martingales. On sait que $\sqrt{n}U_n(\beta_0, \tau)$ est asymptotiquement normal et que

$$\mathcal{I}_n(\beta_0, \tau)^{-1} \int_0^\tau W_n(\beta_n^*, s) S_n^{(0)}(\beta_0, s) \lambda_0(s) ds$$

converge complètement vers

$$\mathcal{I}(\beta_0, \tau)^{-1} \int_0^\tau w(\beta_0, s) s^{(0)}(\beta_0, s) \lambda_0(s) ds,$$

par le corollaire 3.2.1 et le fait que $W_n(\beta_n^*, s)$ converge complètement vers $w(\beta_0, s)$, uniformément sur $[0, \tau]$ (ce résultat n'est pas détaillé ici : les calculs sont en tout point semblables à ceux des chapitres précédents). En outre, par définition de $U_n(\beta_0, \tau)$, on a la formulation suivante pour le terme principal de (5.3) :

$$\begin{aligned} & \sqrt{n}U_n(\beta_0, \tau) \mathcal{I}(\beta_0, \tau)^{-1} \int_0^\tau w(\beta_0, s) s^{(0)}(\beta_0, s) \lambda_0(s) ds \\ &= g(\beta_0, \tau) \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau (Z_i(u) - E_n(\beta_0, u)) dM_i(u), \end{aligned} \quad (5.5)$$

où

$$g(\beta_0, \tau) = \mathcal{I}(\beta_0, \tau)^{-1} \int_0^\tau w(\beta_0, s) s^{(0)}(\beta_0, s) \lambda_0(s) ds$$

est une constante.

De plus, on a :

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau V_n(\hat{\beta}_n, s) dM_i(s) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau (V_n(\hat{\beta}_n, s) - v(\beta_0, s)) dM_i(s) \quad (5.6)$$

$$+ \frac{1}{\sqrt{n}} \int_0^\tau v(\beta_0, s) \sum_{i=1}^n dM_i(s). \quad (5.7)$$

On sait que (5.6) converge en probabilité vers 0 ; il suffit pour voir cela de constater par exemple que ses variations prévisibles sont majorées par

$$C \left(\sup_{s \in [0, \tau]} |V_n(\hat{\beta}_n, s) - v(\beta_0, s)| \right)^2,$$

avec C une constante positive. Ce terme converge complètement vers 0, ce qui permet d'achever, *via* l'inégalité de Lengart, la démonstration de ce lemme. \square

5.3 Normalité asymptotique de $\sqrt{N_d} - \sqrt{n_d}$

Théorème 5.3.1. *Sous les hypothèses (H1), (H3bis) et (H4-H6), on a :*

- (A) $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$ converge en loi vers une gaussienne de moyenne nulle et de variance $\gamma^2(\beta_0, \tau)$, lorsque $n \rightarrow \infty$;
- (B) Pour toute suite régulière de règles d'arrêt $(T_n)_{n \geq 1}$, on a :

$$\sqrt{T_n}(\mathcal{I}_{T_n}(\hat{\beta}_{T_n}, \tau) - \mathcal{I}(\beta_0, \tau)) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \gamma^2(\beta_0, \tau)), \quad (5.8)$$

lorsque $n \rightarrow \infty$.

De ce théorème découle le résultat principal de ce chapitre.

Théorème 5.3.2. *Sous les hypothèses (H1), (H3bis) et (H4-H6), on a :*

$$\sqrt{N_d} - \sqrt{n_d} \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\gamma^2(\beta_0, \tau)}{4\mathcal{I}(\beta_0, \tau)^3}\right), \quad (5.9)$$

lorsque $d \rightarrow 0^+$.

Démonstration du théorème 5.3.2 : D'après la proposition 2.1.1, nous savons que $(N_d)_{d>0}$ est régulière. De plus, par le corollaire 4.1.1, nous savons que $\hat{\sigma}_n^2 = (\mathcal{I}_n(\hat{\beta}_n, \tau))^{-1}$ converge complètement vers $\sigma^2 = (\mathcal{I}(\beta_0, \tau))^{-1}$, lorsque $n \rightarrow +\infty$. D'après la proposition 5.1.1, il nous suffit, pour achever la démonstration, de démontrer que

$$\sqrt{n_d}(\hat{\sigma}_{N_d}^2 - \sigma^2) \xrightarrow{\mathcal{D}} \mathcal{N}\left(0, \frac{\gamma^2(\beta_0, \tau)}{\mathcal{I}(\beta_0, \tau)^4}\right),$$

et comme

$$\sqrt{n_d}(\hat{\sigma}_{N_d}^2 - \sigma^2) = \sqrt{\frac{n_d}{N_d}} \frac{1}{\mathcal{I}(\beta_0, \tau)\mathcal{I}_{N_d}(\hat{\beta}_{N_d}, \tau)} \sqrt{N_d}(\mathcal{I}_{N_d}(\hat{\beta}_{N_d}, \tau) - \mathcal{I}(\beta_0, \tau)),$$

par le résultat (B) du théorème 5.3.1, le résultat s'ensuit si $\mathcal{I}_{N_d}(\hat{\beta}_{N_d}, \tau)$ converge en probabilité vers $\mathcal{I}(\beta_0, \tau)$, lorsque $d \rightarrow 0^+$, ce qui est une conséquence du lemme 3.2.1. \square

5.3.1 Démonstration du théorème 5.3.1

Montrons (A). Dans l'expression donnée par le lemme 5.2.2, les θ_k , $k = 0, 1, 2$, sont des termes déterministes, et M_i est une martingale, ce qui permet de déduire que :

$$\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n) - \mathcal{I}(\beta_0)) = o_P(1) + \frac{1}{\sqrt{n}} \sum_{i=1}^n H_i(\beta_0, \tau),$$

où les $H_i(\beta_0, \tau)$ sont des variables aléatoires d'espérances nulles. On pose $\gamma^2(\beta_0, \tau) = \mathbb{E}H_i(\beta_0, \tau)^2$ strictement positive par (H6) et par le théorème de la limite centrale, on

obtient la normalité asymptotique de $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$ vers une loi normale centrée et de variance $\gamma^2(\beta_0, \tau) > 0$, supposée non dégénérée.

Montrons (B). Par la proposition 4.2.1, il suffit de démontrer que $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$ est asymptotiquement équivalent, au sens de la convergence complète, à une suite satisfaisant la condition d'Anscombe et ayant pour distribution asymptotique une loi normale centrée, de variance $\gamma^2(\beta_0, \tau)$. Le lemme 5.2.1 nous donne un équivalent asymptotique (au sens de la convergence complète) de $\sqrt{n}(\mathcal{I}_n(\hat{\beta}_n, \tau) - \mathcal{I}(\beta_0, \tau))$, qui est :

$$\varepsilon_n(\tau) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau V_n(\hat{\beta}_n, s) dM_i(s) \quad (5.10)$$

$$+ (\mathcal{I}(\beta_0, \tau))^{-1} \sum_{i=1}^n \int_0^\tau W_n(\beta_n^*, s) S_n^{(0)}(\beta_0, s) \lambda_0(s) ds \sqrt{n} U_n(\beta_0, \tau) \quad (5.11)$$

$$+ \sqrt{n} \sum_{k=0}^2 \int_0^\tau \theta_k(\beta_0, s) (S_n^{(k)}(\beta_0, s) - s^{(k)}(\beta_0, s)) \lambda_0(s) ds, \quad (5.12)$$

où β_n^* appartient au segment d'extrémités β_0 et $\hat{\beta}_n$ (donc converge complètement vers β_0). De manière évidente, la continuité uniforme en probabilité est stable pour la sommation, aussi suffit-il de démontrer la condition d'Anscombe pour chaque terme (5.10)-(5.12).

Traitons dans un premier temps le terme (5.10). On démontre l'équivalence asymptotique complète de

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau V_n(\hat{\beta}_n, s) dM_i(s)$$

et de

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau v(\beta_0, s) dM_i(s),$$

ce qui nous permet d'obtenir la convergence complète vers 0 de

$$\sup_{s \in [0, \tau]} \left| V_n(\hat{\beta}_n, s) - v(\beta_0, s) \right|$$

puis en utilisant l'inégalité maximale pour martingale comme cela a déjà été fait pour démontrer le lemme 4.2.3. Comme les variables aléatoires $\int_0^\tau v(\beta_0, s) dM_i(s)$ sont centrées, indépendantes et identiquement distribuées, on conclut par le lemme 5.1.1.

Concernant le terme (5.3), il est facile d'établir que

$$(\mathcal{I}(\beta_0, \tau))^{-1} \sum_{i=1}^n \int_0^\tau W_n(\beta_n^*, s) S_n^{(0)}(\beta_0, s) \lambda_0(s) ds \xrightarrow{c} g(\beta_0, \tau),$$

lorsque n tend vers l'infini. De plus, nous avons vu au chapitre 4 que $\sqrt{n} U_n(\beta_0, \tau)$ est asymptotiquement équivalent à $\sqrt{n} \tilde{U}_n(\beta_0, \tau)$ qui satisfait la condition d'Anscombe, on conclut alors par le lemme 5.1.2.

Le terme (5.12) s'écrit sous la forme suivante

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n P_i,$$

où les $P_i, 1 \leq i \leq n$, sont des variables aléatoires indépendantes et identiquement distribuées. Puisque l'on a $s^{(2)}(\beta_0, s) = \mathbb{E}S_n^{(2)}(\beta_0, s)$, on peut écrire (5.12) de la façon suivante :

$$\begin{aligned} & \sqrt{n} \sum_{k=0}^2 \int_0^\tau \theta_k(\beta_0, s) (S_n^{(k)}(\beta_0, s) - s^{(k)}(\beta_0, s)) \lambda_0(s) ds \\ &= \sum_{k=0}^2 \frac{1}{\sqrt{n}} \sum_{i=1}^n \int_0^\tau \theta_k(\beta_0, s) (Z_i(s)^k e^{\beta_0 Z_i(s)} Y_i - \mathbb{E}Z_i(s)^k e^{\beta_0 Z_i(s)} Y_i) \lambda_0(s) ds \\ &= \sum_{k=0}^2 \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(\int_0^\tau P_i^{(k)}(s) ds - \mathbb{E} \int_0^\tau P_i^{(k)}(s) ds \right), \end{aligned}$$

avec $P_i^{(k)}(s) = Z_i(s)^k \exp(\beta_0 Z_i(s)) Y_i(s) \lambda_0(s)$, et on a bien $\int_0^\tau P_i^{(k)}(s) ds$ indépendants et identiquement distribués. Cela nous permet alors d'appliquer le lemme 5.1.2 pour constater que (5.12) vérifie la condition d'Anscombe.

Enfin, par les lemmes 5.2.1 et 5.2.2 et le résultat (A) du théorème, $\varepsilon_n(\tau)$ est asymptotiquement normale, centrée, de variance asymptotique égale $\gamma^2(\beta_0, \tau)$, ce qui termine la preuve. \square

Chapitre 6

Cadre multidimensionnel

Sommaire

6.1	Introductions et hypothèses	97
6.2	Démonstration du théorème	99
6.3	Vérification des conditions (C-I)-(C-IV) pour le modèle de Cox	100

omme précisé auparavant, nous avons restreint l'étude, au cours des chapitres 3 et 4, au cas d'un paramètre de régression β_0 réel et par conséquent considéré une seule variable explicative Z , elle aussi réelle. Il est néanmoins possible de généraliser l'ensemble des démonstrations et résultats au cas multidimensionnel. Un élément essentiel est apporté par la continuité de la fonction déterminant, qui nous permet ainsi de considérer des suites de matrices définies positives.

Nous verrons dans un premier temps un exemple de règle d'arrêt visant à minimiser le plus grand demi-axe d'une ellipsoïde de confiance, puis nous donnerons les adaptations nécessaires à l'extension des résultats des chapitres 3 et 4. Dans ce chapitre, $\|\cdot\|$ désigne la norme euclidienne sur \mathbb{R}^p , ou une norme matricielle sur $\mathcal{M}_p(\mathbb{R})$, selon le contexte.

6.1 Introductions et hypothèses

ésormais le paramètre d'intérêt β_0 varie dans un espace de dimension p , avec $p > 1$. Nous conservons les mêmes notations que précédemment pour U_n, \mathcal{I}_n et C_n (déjà définis dans un cadre multidimensionnel) et dans le cas du modèle à risques proportionnels de Cox, l'article [AG82] nous permet d'obtenir la convergence suivante :

$$\left(\hat{\beta}_n - \beta_0\right)^T n\mathcal{I}_n(\hat{\beta}_n) \left(\hat{\beta}_n - \beta_0\right) \xrightarrow{\mathcal{D}} \chi_p^2, \quad n \rightarrow +\infty, \quad (6.1)$$

où $\hat{\beta}_n$ désigne l'estimateur du paramètre β_0 pour un échantillon de taille n et χ_p^2 désigne la loi du chi-deux à p degrés de liberté. Dans toute la suite, on supposera (6.1) vérifiée (elle l'est dans le cas semi-paramétrique du modèle de Cox, il suffit par exemple de consulter [ABGK93] ou encore [AG82]). Il paraît alors naturel de considérer pour région de confiance au seuil $\alpha \in]0, 1[$:

$$RC_n(\alpha) = \left\{ \beta \in \mathbb{R}^p; \left(\hat{\beta}_n - \beta\right)^T n\mathcal{I}_n(\hat{\beta}_n) \left(\hat{\beta}_n - \beta\right) \leq \chi_{p,1-\alpha}^2 \right\},$$

par un raisonnement similaire à celui de l'article de [DG00]. Ici, $\chi_{p,1-\alpha}^2$ désigne le quantile d'ordre $1 - \alpha$ d'une loi du chi-deux à p degrés de liberté. Il est possible de définir d'autres règles d'arrêt (voir par exemple [GMS97]). Il nous reste maintenant à établir une règle d'arrêt qui soit pertinente, montrer sa régularité et enfin généraliser les théorèmes montrés dans le cas unidimensionnel des chapitres 3 et 4 à ce cas.

Considérons l'ellipsoïde définie par :

$$\left(\hat{\beta}_n - \beta_0\right)^T n\mathcal{I}_n(\hat{\beta}_n) \left(\hat{\beta}_n - \beta_0\right) = \chi_{p,1-\alpha}^2.$$

On peut montrer que la longueur du plus grand axe d'une telle ellipsoïde vaut :

$$\frac{2}{\sqrt{\Phi_{\min} \left(\frac{n\mathcal{I}_n(\hat{\beta}_n)}{\chi_{p,1-\alpha}^2} \right)}},$$

où Φ_{\min} désigne la fonction donnant la valeur propre minimale. Pour imposer que chaque composante du vecteur β_0 soit dans un intervalle de confiance de longueur au plus $2d$, il suffit d'imposer :

$$n \geq \frac{\chi_{p,1-\alpha}^2}{d^2 \Phi_{\min} \left(\mathcal{I}_n(\hat{\beta}_n) \right)}.$$

On choisira alors pour variable d'arrêt la variable N_d , définie par :

$$N_d = \min \left\{ n \geq n_0; n \geq \frac{\chi_{p,1-\alpha}^2}{d^2 \Phi_{\min}^* \left(\mathcal{I}_n(\hat{\beta}_n, \cdot) \right)} \right\},$$

où

$$\Phi_{\min}^* \left(\mathcal{I}_n(\hat{\beta}_n) \right) = \min \left\{ \Phi_{\min} \left(\mathcal{I}_n(\hat{\beta}_n) \right), \varepsilon'_n \right\},$$

avec $(\varepsilon'_n)_{n \geq 1}$, une suite de nombres réels positifs tendant vers l'infini.

Ces quelques considérations nous fournissent alors la définition de la nouvelle variable d'arrêt :

$$N_d = N_{d, n_0, \alpha} := \min \left\{ n \geq n_0; \Phi_{\min}(\mathcal{I}_n(\hat{\beta}_n))^* \geq \frac{\chi_{p, 1-\alpha}^2}{d^2 n} \right\}, \quad (6.2)$$

où α et n_0 seront fixés dans toute la suite et donc omis dans les notations. Ainsi N_d est une règle régulière d'arrêt, dans le sens où

$$\frac{N_d}{n_d} \xrightarrow{d \rightarrow 0^+} 1, \quad (6.3)$$

avec $n_d = \min \left\{ n \geq 1; \Phi_{\min}(\mathcal{I}(\beta_0)) \geq \frac{\chi_{p, 1-\alpha}^2}{nd^2} \right\}$ (les démonstrations de ces propriétés ne nécessitent pas d'adaptation particulière, puisque l'on est ramené à une inégalité réelle, et l'on peut donc mener des raisonnements similaires à ceux du chapitre 2, le lecteur intéressé pourra se référer à [GMS97, p.329]).

Théorème 6.1.1. *On suppose que*

C-I

$$\frac{\|U_n(\beta_0)\|}{\Phi_{\min}(\mathbf{IE}\mathcal{I}_n(\beta_0; \tau))} \xrightarrow{c} 0,$$

lorsque $n \rightarrow \infty$.

C-II $\beta \mapsto \mathcal{I}_n(\beta, t)$ est une fonction continue sur \mathcal{B} , uniformément sur $[0, \tau]$, et l'on suppose que, pour $\varepsilon > 0$, il existe $\delta_\varepsilon > 0$ tel que

$$\sum_{n \geq 1} \mathbb{P} \left(\sup_{\beta \in B(\beta_0; \delta_\varepsilon)} \sup_{t \in T} \|\mathcal{I}_n(\beta, t) - \mathbf{IE}\mathcal{I}_n(\beta_0, t)\| > \varepsilon \right) < \infty;$$

$\mathbf{IE}\mathcal{I}_n(\beta_0) \rightarrow \mathcal{I}(\beta_0)$, matrice définie positive, lorsque $n \rightarrow +\infty$.

C-III $\sqrt{n}U_n(\beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0))$, lorsque $n \rightarrow +\infty$.

C-IV Pour toute suite $(T_n)_{n \geq 1}$ de variables aléatoires régulière, on a :

$$\sqrt{T_n}U_{T_n}(\beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)).$$

C-V Les matrices $\mathcal{I}_n(\beta)$ sont presque sûrement définies positives.

Alors $T_n(\hat{\beta}_{T_n} - \beta_0)^T \mathcal{I}_{T_n}(\hat{\beta}_{T_n})(\hat{\beta}_{T_n} - \beta_0) \rightarrow \chi_p^2$.

Pour montrer ce résultat, il nous suffit d'une part de généraliser la démonstration du théorème 3.1.1 au cas multidimensionnel, et d'autre part, de vérifier les conditions **(C-I)**-**(C-IV)**. L'un des points clé de ces démonstrations résulte de la continuité du déterminant, par le biais du lemme suivant.

Lemme 6.1.1. *Soit $(X_n)_{n \geq 1}$, une suite de matrices semi-définies positives, qui converge vers X , matrice définie positive. Alors il existe un entier n_0 tel que $(X_n)_{n \geq n_0}$ soit une suite de matrices définies positives.*

Démonstration : Considérons l'image de la suite $(X_n)_{n \geq 1}$ par l'application déterminant définie par :

$$\begin{cases} \mathcal{M}_p(\mathbb{R}) & \rightarrow \mathbb{R}, \\ X & \mapsto \det X, \end{cases}$$

où $\mathcal{M}_p(\mathbb{R})$ désigne les matrices $p \times p$ à coefficients réels.

Cette application étant continue, on sait que la suite $(\det(X_n))_{n \geq 1}$ converge vers $\det X$, réel strictement positif, elle est donc strictement positive à partir d'un certain rang, ce qui achève la démonstration du lemme. \square

Remarque 6.1.1. *On sait que les matrices $\mathcal{I}_n(\beta, t)$ sont semi-définies positives [AG82, p.1103].*

6.2 Démonstration du théorème

Reprenons pas à pas la démonstration du théorème 3.1.1 et attardons-nous sur les quelques différences qui résultent du cadre multidimensionnel, les grandes lignes de la démonstrations étant les mêmes.

Considérant, pour $n \geq 1$ et $\delta > 0$, l'ensemble

$$A_n = \left\{ \|U_n(\beta)\| \leq \frac{\delta}{2} \inf_{\beta \in B(\beta_0; \delta)} \mathcal{I}_n(\beta) \right\},$$

on a le fait que

$$A_n \subset \left\{ \|\hat{\beta}_n - \beta_0\| \leq \delta \right\}.$$

En effet, par concavité de la log-vraisemblance [AG82], on a (voir [DG00])

$$A_n \subset \{ \forall \beta \in \partial B(\beta_0; \delta), C_n(\beta) - C_n(\beta_0) \leq 0 \}, \quad (6.4)$$

ce qui, à l'aide du théorème des accroissements finis, nous donne

$$C_n(\beta) - C_n(\beta_0) = \delta u^T U_n(\beta_0) - \frac{\delta^2}{2} u^T \mathcal{I}_n(\tilde{\beta}) u,$$

où $\tilde{\beta}$ appartient au segment d'extrémités β_0 et β , et $u = (\beta - \beta_0) / \|\beta - \beta_0\|$.

En utilisant le fait que $\Phi_{\min}(\mathcal{I}_n(\tilde{\beta})) \leq u^T \mathcal{I}_n(\tilde{\beta}) u$, on obtient (6.4). Considérons désormais l'ensemble B_n défini par

$$B_n = \left\{ \sup_{\beta \in B(\beta_0, \delta)} \|\mathcal{I}(\beta) - \mathbb{E}\mathcal{I}_n(\beta_0)\| \leq C \Phi_{\min}(\mathbb{E}\mathcal{I}_n(\beta_0)) \right\},$$

où $C \in]0, 1]$; il nous est possible de reprendre le squelette de la démonstration du théorème 3.1.1 et de nous ramener à montrer séparément que les séries de termes généraux $\mathbb{P}(B_n \cap A_n^c)$ et $\mathbb{P}(B_n^c)$ sont convergentes.

Cette première convergence se déduit directement de la condition **(C-II)**. En effet, par lemme 6.1.1, on sait que la suite $(\mathbb{E}\mathcal{I}_n(\beta_0))_{n \geq 1}$ est une suite de matrices définies positives à partir d'un certain rang, et donc que la suite $(\Phi_{\min}(\mathbb{E}\mathcal{I}_n(\beta_0)))_{n \geq 1}$ tend vers un réel strictement positif, il est donc possible de la minorer par $\eta > 0$, à partir d'un certain rang, et de conclure par **(C-II)**.

La seconde convergence nécessite plus d'adaptation à ce cadre multidimensionnel. Par définitions de A_n^c et de B_n , on a

$$\|\mathcal{I}_n(\beta) - \mathbb{E}\mathcal{I}_n(\beta_0)\| \leq C\Phi_{\min}(\mathbb{E}\mathcal{I}_n(\beta_0)).$$

Pour conclure, on applique le lemme 6.1.1, en remarquant que par **(C-II)** et l'hypothèse **(C-V)**, $(\mathbb{E}\mathcal{I}_n(\beta_0))_{n \geq 1}$ est une suite de matrices définies positives à partir d'un certain rang, et que le fait que $(\mathcal{I}_n(\beta))_{n \geq 1}$ soit semi-définie positive suffit à montrer le lemme B.2.1.

On obtient alors :

$$\begin{aligned} \Phi_{\min}(\mathcal{I}_n(\beta)) &\geq \Phi_{\min}(\mathbb{E}\mathcal{I}_n(\beta_0)) - \|\mathcal{I}_n(\beta) - \mathbb{E}\mathcal{I}_n(\beta_0)\| \\ &\geq (C + 1)\Phi_{\min}(\mathbb{E}\mathcal{I}_n(\beta_0)), \end{aligned}$$

d'où

$$\|U_n(\beta_0)\| \geq \frac{\delta}{2}(C + 1)\Phi_{\min}(\mathbb{E}\mathcal{I}_n(\beta_0)),$$

et par la condition **(C-I)**, on obtient la convergence souhaitée. \square

On désire maintenant établir un résultat analogue à celui du théorème 3.2.1. Nous avons déjà remarqué lors de leurs démonstrations (voir remarque 3.2.2) que le lemme 3.2.1 ainsi que le théorème 3.2.1 se généralisent immédiatement au cas où $\beta_0 \in \mathbb{R}^p$. Cela achève cette partie, puisque ces deux résultats suffisent à obtenir un théorème de convergence en loi.

6.3 Vérification des conditions **(C-I)**-**(C-IV)** pour le modèle de Cox

Formulons dans un premier temps les hypothèses relatives au modèle de Cox. Nous examinons ensuite les conditions **(C-I)** à **(C-V)** et nous verrons que la plupart d'entre elles se déduit directement des résultats établis au cours du chapitre 4.

Hypothèses 4.

$$\mathbf{H-I} \quad \int_0^\tau \lambda_0(t)dt \text{ est fini};$$

H-IIbis Il existe un voisinage \mathcal{B} de β_0 , des fonctions $s^{(0)}$, $s^{(1)}$ et $s^{(2)}$ définies sur $\mathcal{B} \times [0, \tau]$ et à valeurs dans \mathbb{R} , \mathbb{R}^p et $\mathcal{M}_p(\mathbb{R})$ respectivement, telles que pour $j = 0, 1, 2$:

$$\sup_{t \in [0, \tau], \beta \in \mathcal{B}} \|\mathbb{E}S_n^{(j)}(\beta, t) - s^{(j)}(\beta, t)\| \rightarrow 0;$$

H-III En conservant les notations précédentes, on définit

$$e = \frac{s^{(1)}}{s^{(0)}} \text{ et } v = \frac{s^{(2)}}{s^{(0)}} - e^{\otimes 2}.$$

Pour β dans \mathcal{B} , $t \in [0, \tau]$,

$$s^{(1)}(\beta, t) = \frac{\partial}{\partial \beta} s^{(0)}(\beta, t), s^{(2)}(\beta, t) = \frac{\partial^2}{\partial \beta \partial \beta^T} s^{(0)}(\beta, t),$$

$s^{(0)}(\cdot, t)$, $s^{(1)}(\cdot, t)$, et $s^{(2)}(\cdot, t)$ sont des fonctions continues sur \mathcal{B} , uniformément en $t \in [0, \tau]$; $s^{(0)}$ est strictement positive sur $\mathcal{B} \times [0, \tau]$ et

$$\mathcal{I}(\beta_0) = \int_0^\tau v(\beta_0, t) s^{(0)}(\beta_0, t) \lambda_0(t) dt$$

est une matrice définie positive;

H-IV $\{Z_i^{(j)}(t); t \in [0, \tau]\}$, pour i entier naturel et $1 \leq j \leq p$, est à variations bornées, uniformément en $i \geq 1$, i.e.

$$\forall i \geq 1, \int_0^\tau |Z_i^{(j)}(ds)| \leq B < \infty,$$

où B désigne une constante positive.

Montrons une à une les conditions **(C-I)**-**(C-V)**.

Condition (C-I)

Il nous suffit de remarquer que, pour $n \geq 1$, on a

$$U_n(\beta_0) = \begin{pmatrix} U_n^{(1)}(\beta_0) \\ \vdots \\ U_n^{(p)}(\beta_0) \end{pmatrix},$$

où chaque terme $U_n^{(i)}$, $1 \leq i \leq p$, est de la même forme que celui considéré dans le théorème 3.1.1 et vérifie les propriétés suffisantes pour aboutir à sa convergence complète. Ce premier point nous apporte la convergence complète de $U_n(\beta_0)$ vers 0.

En outre, on sait que $(\Phi_{\min}(\mathbb{E}\mathcal{I}_n(\beta_0)))_{n \geq 1}$ est une suite bornée inférieurement à partir d'un certain rang, ce qui nous permet de conclure.

Condition (C-II)

Si l'on reprend la démonstration initiale de (C1), il apparaît suffisant de montrer que

$$\sup_{\beta \in \mathcal{B}} \sup_{s \in [0, \tau]} \|V_n(\beta, s) - v(\beta, s)\|$$

converge complètement vers 0 pour montrer que (C-II) est satisfaite. Par définition de V_n , pour $t \in [0, \tau]$, on a

$$V_n(\beta_0, t) = \frac{S_n^{(2)}(\beta_0, t)}{S_n^{(0)}(\beta_0, t)} - E_n(\beta_0, t)^{\otimes 2}.$$

Si l'on considère chacune des composantes de $S^{(2)}$, on constate avec le raisonnement des chapitres précédents, que $S_n^{(2)}(\beta_0, t)[i, j] - \mathbb{E}S_n^{(2)}(\beta_0, t)[i, j]$ converge complètement vers 0, ce qui nous donne, avec la condition (H-IIbis), la convergence complète du premier terme de V_n .

Pour le second terme de V_n , on écrit par l'inégalité de Cauchy-Schwarz, pour β et t ,

$$\begin{aligned} \|E_n(\beta, t)^{\otimes 2} - \mathbb{E}E_n(\beta, t)^{\otimes 2}\| &\leq \|E_n(\beta, t)\| \|E_n(\beta, t)^T - \mathbb{E}E_n(\beta, t)^T\| \\ &+ \|E_n(\beta, t) - \mathbb{E}E_n(\beta, t)\| \|\mathbb{E}E_n(\beta, t)\|, \end{aligned}$$

et l'on conclut à l'aide des inégalités et raisonnements du chapitre 4 appliqués à chaque composante des vecteurs ou matrices considérés.

Condition (C-III) voir [AG82, démonstration du théorème 3.2 p.1106]

Condition (C-IV) Il est aisé de généraliser les démonstrations effectuées au cours du chapitre 4 pour démontrer (C1). En effet, on considère la version multivariée du théorème de Rebolledo pour la convergence en loi, et l'on remarque que la convergence en probabilité, presque sûre et complète dans \mathbb{R}^p est équivalente à celle de chaque composante.

En ce qui concerne la démonstration de la condition d'Anscombe pour le version multidimensionnelle de \tilde{U}_n (chapitre 4, page 75) \tilde{U}_n , il est possible de reprendre des démonstrations identiques, car la démonstration de l'équivalence pour la convergence complète ne fait intervenir que des vecteurs de \mathbb{R}^p , donc des considérations semblables à celles utilisées pour démontrer (C-I) nous permettent de conclure. Il reste à voir que \tilde{U}_n est asymptotiquement normale, ce qui est automatiquement obtenu lors de l'utilisation du théorème de Rebolledo multidimensionnel.

Condition (C-V) Elle nous est donnée par le lemme 6.1.1, et par la remarque 6.1.1.

Toutes ces considérations nous permettent ainsi d'étendre les résultats des chapitres 4 et 6 au cadre multidimensionnel et de justifier de la cohérence des procédures d'arrêt dans le cas de plusieurs variables explicatives.

Chapitre 7

Simulations numériques

Sommaire

7.1	Simulations et méthodes utilisées	105
7.2	Observations et résultats	105

ous nous intéressons désormais à des considérations plus pratiques qui nous permettent de vérifier les résultats précédemment montrés, dans le cas d'une variable aléatoire réelle. Les outils utilisés sont Matlab et R (plus particulièrement la fonction de calcul de l'estimateur du paramètre de régression pour le modèle de Cox).

La préoccupation principale de ce chapitre consiste essentiellement à vérifier les résultats asymptotiques obtenus, c'est pourquoi nous ne traiterons pas les cas où la taille initiale de l'échantillon est insuffisante. Nous effectuerons des comparaisons de résultats pour différentes valeurs de β_0 , du taux de censure, ou encore de d .

7.1 Simulations et méthodes utilisées

ette partie vise à explorer qualitativement les propriétés précédemment obtenues pour l'estimation séquentielle du paramètre de régression du modèle de Cox, dans le cas particulier d'une règle séquentielle basée sur des intervalles de confiance. Ainsi, notre premier travail consiste à simuler les durées de vie censurées. Il nécessite plusieurs choix, notamment celui des lois suivies par la variable de censure ou encore par les variables explicatives. Dans toute cette partie, nous nous restreindrons au cas où $p = 1$. Par ailleurs, le niveau de confiance sera fixé à 5%.

Dans un premier temps, le programme simule les données d'un échantillon de taille N formé de $(X_i, \delta_i, Z_i)_{1 \leq i \leq N}$, avec les mêmes notations que celles du chapitre 1. On choisit la loi de Weibull pour modéliser la censure (il est possible d'ajuster les paramètres pour faire varier le taux de censure) ainsi que la loi de la fonction de risque de base. Cela se justifie à la fois par la diversité des densités possibles pour une loi de Weibull en faisant varier les paramètres, mais aussi par la simplicité de simulation des variables aléatoires issues de ces lois. Enfin, les variables explicatives suivront une loi uniforme sur l'intervalle $[0, 1]$.

Par mesure de simplicité, nous réalisons dans un premier temps un échantillon de taille N suffisante fournie par l'utilisateur. Il sera ensuite tronqué pour que sa taille corresponde à la taille initiale n_0 . Un défaut qui résulte de ce choix d'implémentation réside dans le fait qu'il est difficile pour l'utilisateur de prévoir la taille de l'échantillon maximale requise. Rappelons cependant qu'il s'agit ici de simulations, par conséquent ce choix a pour but d'éviter de temps de calcul trop élevés (d'autant que le temps nécessaire à une simulation peut se révéler considérable pour des tailles d'échantillons importantes). Il est également possible de décider d'un « pas », qui sera donné sous la forme d'un entier positif définissant le nombre de nouvelles observations à ajouter à l'échantillon avant de procéder au nouveau calcul de la règle d'arrêt.

Le plus fréquemment, nous répétons la procédure d'arrêt pour une centaine d'échantillons simulés indépendamment et effectuons quelques statistiques descriptives sur les résultats obtenus. Par la suite, A désignera le nombre d'échantillons considérés dans l'étude menée.

Le dernier choix incombant à l'utilisateur est bien sûr celui de la règle d'arrêt. Dans un premier temps, cela équivaut simplement à la donnée de d , réel strictement positif, ainsi qu'à celle de l'entier n_0 . Le tableau 7.1 présente un récapitulatif de l'ensemble des paramètres du programme.

7.2 Observations et résultats

n constate immédiatement d'après les tableaux de données obtenus (tableau 7.2) que la taille d'échantillon requise pour une amplitude décroissante de l'intervalle de confiance s'accroît. Ces premières données sont constituées des moyennes des

Notation	Définition
d	amplitude maximale de l'intervalle de confiance
β_0	vraie valeur du paramètre de régression
N	taille totale de l'échantillon simulé
(a_C, b_C)	paramètres de la loi de Weibull pour C
(a_λ, b_λ)	paramètres de la loi de Weibull pour λ_0
(a_Z, b_Z)	paramètres de la loi uniforme pour $Z \sim \mathcal{U}([a_Z, b_Z])$

TAB. 7.1 – Paramètres à fournir par l'utilisateur

d	Moyenne($\hat{\beta}_{N_d}$) (écart-type)	Moyenne(N_d) (écart-type)	I.C.
5	1.0502 (1.1761)	13 (3)	95
2	0.9953 (0.5631)	57 (8)	91
1	1.0413 (0.2605)	218 (19)	95
0.8	1.0175 (0.1702)	339 (21)	99
0.6	1.0128 (0.1627)	589 (26)	94
0.5	1.0000 (0.1458)	849 (31)	93
0.4	1.0009 (0.0950)	1343 (25)	98
0.09	1.0190 (0.0243)	26247 (19)	93
0.07	1.0021 (0.0186)	43174 (258)	94
0.05	1.0011 (0.0132)	84624 (367)	94
0.0027	0.9996 (0.0069)	290222 (724)	98

TAB. 7.2 – Estimations séquentielles de $\beta_0 = 1$ pour différentes valeurs de d , $A = 100$

quantités obtenues pour $\hat{\beta}_{N_d}$ et N_d (deuxième et troisièmes colonnes) pour lesquelles on a pris soin de calculer l'écart-type. La dernière colonne présente le pourcentage d'appartenance de β_0 à l'intervalle de confiance. On peut remarquer qu'il respecte les 95 % attendus. Afin d'illustrer ce comportement, on trace l'évolution de N_d en fonction de d (pour $\beta_0 = 1$ fixé). Les résultats obtenus (figure 7.1) représentent la moyenne de ceux obtenus au cours des A expériences effectuées ($A = 100$). On trace la courbe donnée par

$$f : d \mapsto \frac{u_\alpha^2 \hat{\sigma}_{N_{\max}}^2}{d^2},$$

où $\hat{\sigma}_{N_{\max}}^2$ représente l'estimateur de σ^2 obtenu pour la taille d'échantillon la plus élevée N_{\max} . Il est alors possible de voir graphiquement que les moyennes effectuées sur les A itérations de la procédure séquentielle sont en accord avec la courbe tracée, et donc que les simulations suivent approximativement l'équation $EN_d = f(d)$.

On décide désormais d'observer l'influence du taux de censure sur la qualité de l'estimation. Les résultats numériques sont répertoriés dans le tableau 7.3. Plus l'échantillon est soumis à une forte censure, pour d fixé, plus la taille requise de l'échantillon est grande. Cela correspond au fait que la variance croît avec le taux de censure.

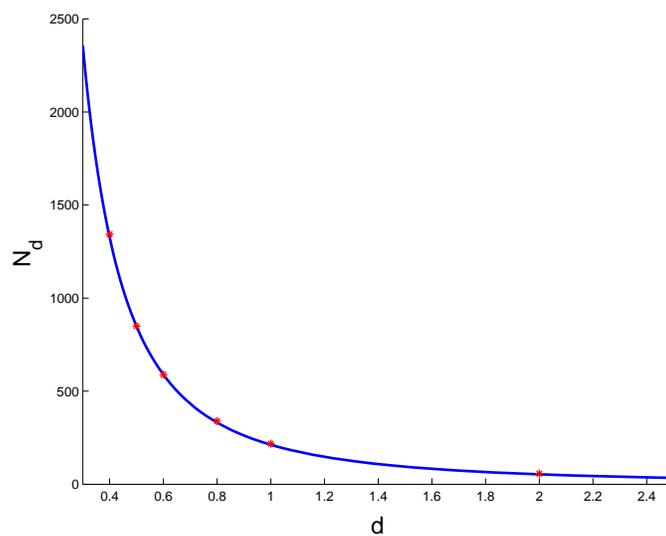


FIG. 7.1 – Evolution de N_d en fonction de d , pour $\beta_0 = 1$ et $A = 100$.

Taux de censure	7 %	30 %	60 %	80 %
Moyenne(N_1)	218(19)	271(22)	453(35)	930(76)
Moyenne($\hat{\beta}_{N_d}$)	1.041(0.26)	1.022(0.26)	1.026(0.27)	0.9810(0.26)
Méthode 1	14(3.88)	15.19(3.90)	33.24(5.77)	60.71(7.79)
Méthode 2	14(0.09)	17.45(0.06)	29.26(0.03)	59.94(10.11)

TAB. 7.3 – Impact du taux de censure

Pour un faible taux de censure (7 %), les conditions asymptotiques sont atteintes pour des tailles relativement faibles d'échantillon, puisqu'environ 220 observations suffisent à obtenir un intervalle de confiance de longueur 1 pour $\beta_0 = 1$ (se référer au tableau 7.3). Il nous est en outre possible d'estimer la variance asymptotique $\mathcal{I}(\beta_0)^{-1}$ par deux méthodes. La première consiste à calculer la moyenne des estimations des A variances pour les A échantillons considérés (elle est intitulée « Méthode 1 » dans les tableaux et figures). La seconde consiste à calculer la variance empirique des A observations de $\sqrt{N_d}(\hat{\beta}_{N_d} - \beta_0)$, dont on a montré qu'elle convergeait vers $\mathcal{I}(\beta_0)^{-1}$ (« Méthode 2 »). Son calcul nécessite de connaître la valeur du paramètre β_0 et il est donc impossible de calculer cette quantité dans les cas pratiques. Elle nous permet néanmoins d'obtenir un bon critère de comparaison, puisque la variance asymptotique est inconnue.

La variance asymptotique estimée croît avec le taux de censure, ce qui s'explique par un fort taux de variabilité des variances obtenues.

Pour d'autres taux de censure, aussi bien que pour d'autres valeurs de β_0 , il est possible de constater que $d = 1$ donne de bons résultats. Nous souhaitons illustrer ces propos par une courte étude de l'influence du choix du paramètre n_0 . Pour ce faire, nous fixons le taux de censure à 30%. Il est effectivement possible d'ajuster les paramètres de la loi de Weibull qui modélise la censure pour obtenir le taux de censure souhaité ; notons néanmoins que ce réglage dépend de la valeur choisie pour β_0 . Considérant une taille d'échantillon de 10 comme insuffisante pour obtenir un résultat pertinent pour l'intervalle de confiance correspondant, on décide d'observer le comportement de l'amplitude de l'intervalle de confiance obtenu dans de telles conditions.

Dans cette série de simulations, on prend $\hat{\sigma}_n^* = \max(10^{-9}, \hat{\sigma}_n)$, mais les valeurs nulles de $\hat{\sigma}_n$ se traduisent par de fortes amplitudes observées que nous excluons lors du calcul de la moyenne des amplitudes. Nous sommes en effet intéressés par des cas « litigieux » où l'amplitude observée est finie puisqu'ils constituent une limite à notre procédure d'arrêt.

Notre objectif est de considérer les résultats obtenus afin d'en déduire une valeur de d qui permette d'éviter les arrêts prématurés de la procédure d'arrêt.

Les résultats de cette étude sont donnés dans le tableau 7.4, on peut voir que les amplitudes dépassent très fréquemment la valeur 1, qui devient alors notre valeur d'intérêt. Les quartiles nous permettent de juger de la répartition des données obtenues et de constater que si l'on exclut les valeurs trop grandes, l'amplitude se concentre autour de sa médiane.

Le graphe 7.2 présente la répartition des valeurs obtenues et permet de conforter la valeur $d = 1$ comme queue de distribution des amplitudes obtenues.

Par conséquent, nous choisirons $d = 1$ dans toute la suite puisque nous avons montré qu'elle n'impose pas de restriction spécifique (dans le cadre des données simulées), lorsque l'on considère des échantillons de faibles tailles.

On remarque en outre que pour la valeur de $d = 1$, la taille de l'échantillon requis croît avec la valeur absolue de la variable β_0 (voir figure 7.3). Il nous est encore une

β_0	Taux de censure	Moyenne(amplitudes)	quartiles		
			25%	50%	75%
0	28.3%	6.80(2.79)	5.34	6.28	7.53
1	27.8%	6.98(5.35)	5.33	6.34	7.50
10	27.3%	31.82(60.18)	13.89	18.90	29.50
20	24%	1364880(2×10^7)	29.34	45.36	83.67

TAB. 7.4 – Valeurs de l’amplitude de l’intervalle de confiance pour $N = n_0 = 10$

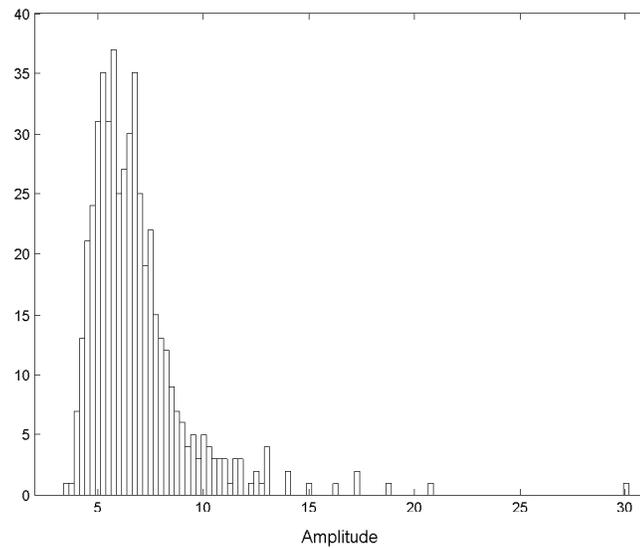


FIG. 7.2 – Effectifs des amplitudes obtenues pour des échantillons de taille 10, pour $\beta_0 = 1$

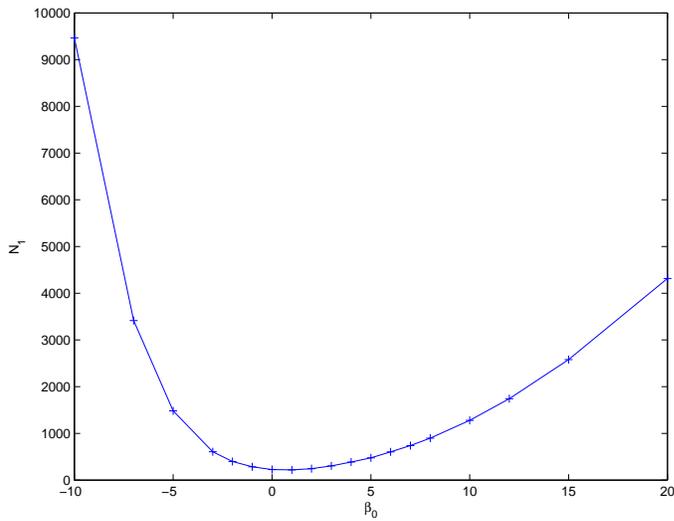


FIG. 7.3 – Evolution de Moyenne(N_d) en fonction de β_0 .

β_0	-5	-1	0	1
Moyenne($\hat{\beta}_{N_d}$)	-5.01(0.25)	-0.94(0.27)	0.09(0.20)	1.03(0.25)
Moyenne(N_1)	564(41)	276(23)	258(21)	272(19)
Méthode 1	34.97	19.61	10.73	16.95
Méthode 2	36.43	17.76	16.63	17.54

TAB. 7.5 – Influence de β_0 sur N_1 , pour un taux de censure de 30% , $A = 100, d = 1$

fois possible d'expliquer ce phénomène par une grande hétérogénéité des données, car elles sont davantage soumises au comportement de la variable explicative. Afin de juger si l'augmentation de N_d avec $|\beta_0|$ est due à une plus forte hétérogénéité des données ou à l'influence du taux de censure (il varie effectivement avec la valeur de β_0), on décide de fixer le taux de censure et de faire varier β_0 . Les résultats sont présentés dans le tableau 7.5. Les résultats sont semblables à ceux obtenus pour un taux de censure variable, ce qui amène à penser que le taux de censure influe peu sur les valeurs de N_1 (la taille requise de l'échantillon est une fois de plus minimale pour de faibles valeurs de $|\beta_0|$), c'est donc effectivement l'hétérogénéité des données qui accroît la valeur de N_1 , lorsque $|\beta_0|$ augmente.

Cela justifie le choix d'un seuil d d'arrêt dépendant de la valeur du paramètre de régression, et permet ainsi un gain de temps considérable, comme le démontre le tableau 7.6. Dans ce tableau, on note le critère d'arrêt $d_n = d + \gamma \hat{\beta}_n$. Le temps est donné en secondes. Pour les valeurs de β_0 de 2 et 5, on a aussi simulé les cas où $\gamma = 0.5$ et $\gamma = 0.2$ afin de pouvoir juger du temps de simulation pour une même précision relative, on constate qu'il est inférieur dans le cas de critère relatif d'arrêt.

Considérons désormais les estimations de la variance asymptotique (voir tableau 7.7 et figure 7.4), pour diverses valeurs du paramètre β_0 , pour d fixé à 1.

β_0	1	1	2	2	2	5	5	5
d	10^{-5}	1	10^{-5}	10^{-5}	1	10^{-5}	10^{-5}	1
γ	1	0	1	0.5	0	1	0.2	0
Temps moyen	19.16	97.52	0.83	29.39	38.3	0.98	89.73	583.71
Moyenne($\hat{\beta}_n$)	1.20	0.93	2.07	2.05	1.95	5.01	5.01	5.04
Moyenne(N_d)	196	216	101	252	245	200	481	484

TAB. 7.6 – Influence du critère relatif d’arrêt sur la durée de la simulation, $A = 100$.

β_0	-10	-7	-5	-3	-2	-1	0	1	2
Moyenne(N_1)	9466	3415	1481	607	398	284	225	218	245
Méthode 1	611	220	85	46	38	18	14	14	16
Méthode 2	550	199	95	39	26	18	14	14	16

β_0	3	4	5	6	7	8	10	12	15	20
Moyenne(N_1)	305	386	477	605	740	897	1278	1743	2579	4316
Méthode 1	20	25	35	31	53	57	80	96	175	316
Méthode 2	20	25	31	39	48	58	83	113	167	280

TAB. 7.7 – N_1 en fonction de β_0 .

Lorsque $|\beta_0|$ augmente, sa variance croît également. On constate que la valeur la plus faible de la variance est obtenue pour $\beta_0 = 1$. Cela est principalement dû à la forme particulière choisie pour la variable explicative dans les simulations (elle suit une loi uniforme sur $[0, 1]$). On peut remarquer que la variance sera minimale en 0 pour une loi uniforme dont l’intervalle est symétrique par rapport à 0, et positive (égale à 1) si l’on prend un $Z \sim \mathcal{U}([a, a + 1])$, $a > 0$. Il apparaît difficile d’expliquer ce phénomène.

Nous souhaitons désormais pouvoir juger de la normalité asymptotique démontrée dans le théorème 3.2.1, cela nécessite une quantité conséquente de données. Nous ferons le choix de prendre $\beta_0 = 1$, puisque cette valeur nous permet de réduire la taille de l’échantillon, et par conséquent de gagner du temps de calcul.

Notre résultat principal est le suivant :

$$\sqrt{N_d}(\hat{\beta}_{N_d} - \beta_0) \xrightarrow{\mathcal{D}} \mathcal{N}(0, \mathcal{I}(\beta_0)^{-1}).$$

On décide d’approcher la loi asymptotique en considérant de nombreuses simulations. En prenant $d = 1$ et $\beta_0 = 1$, on simule ainsi 10000 échantillons pour lesquels on applique la procédure séquentielle. On remarque sur les graphes 7.5 et 7.6 de la fonction de répartition empirique et sur le diagramme empirique que la loi asymptotique considérée est proche de celle d’une loi normale. Il est intéressant de constater que ces valeurs sont correctes pour des tailles d’échantillon relativement faibles.

On peut constater que pour une amplitude relativement faible, il est possible d’obtenir une estimation pertinente de β_0 . La procédure consistant à choisir un cri-

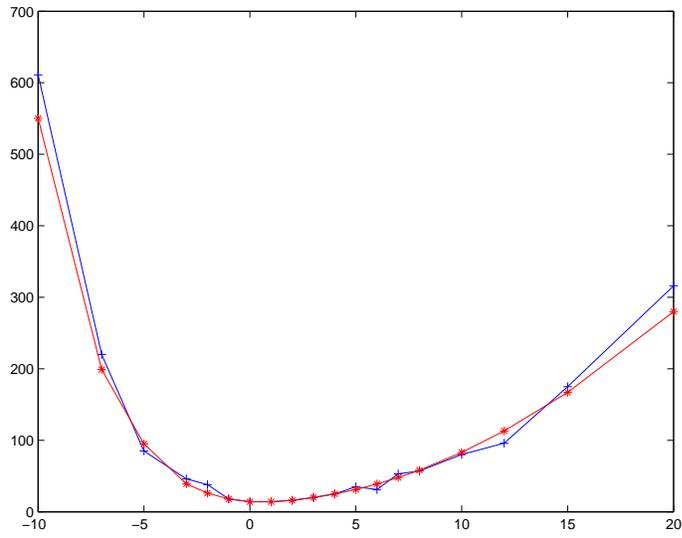


FIG. 7.4 – Estimation de la variance en fonction de β_0 , données du tableau 7.7

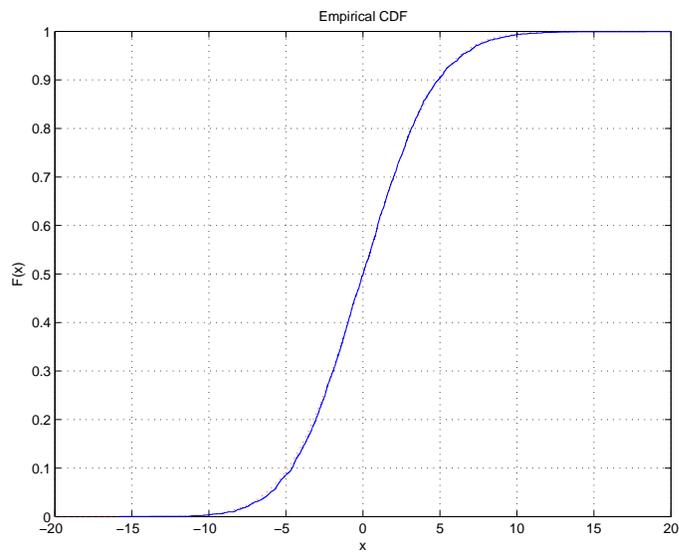


FIG. 7.5 – Fonction de répartition empirique

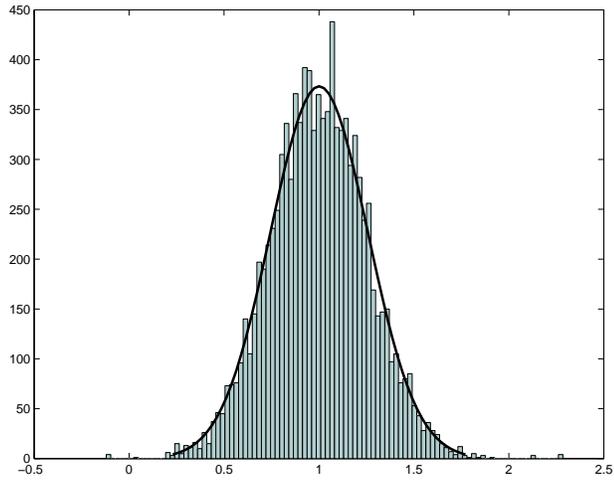


FIG. 7.6 – Comparaison avec la fonction de densité gaussienne

tère relatif pour stopper la procédure peut être utile grâce à sa rapidité et peut ainsi permettre d'obtenir rapidement un ordre de grandeur pour β_0 , tout en ayant une précision finale donnée. Il serait néanmoins intéressant d'étudier le comportement de cette règle d'arrêt pour un modèle soumis à des perturbations afin d'évaluer la robustesse de l'estimateur séquentiel.

Annexe A

Inégalités maximales

ans cette partie, nous nous efforcerons de rassembler tous les résultats utilisés pour démontrer la convergence complète des processus empiriques définis au chapitre 4 en présentant un court résumé des premiers chapitres de [Pol90].

A.1 Définition de la norme d'Orlicz

Définition A.1.1. Soit ϕ une fonction convexe croissante sur \mathbb{R}_+ telle que $0 \leq \phi(0) < 1$. On définit la norme d'Orlicz $\|Z\|_\phi$ d'une variable aléatoire Z par :

$$\|Z\|_\phi = \inf \left\{ C > 0; \mathbb{E}\phi\left(\frac{|Z|}{C}\right) \leq 1 \right\}.$$

Remarque A.1.1.

1. Si l'on prend comme convention $\inf \emptyset = +\infty$, $+\infty$ est une valeur possible pour la borne inférieure ;
2. Si l'on considère $\phi : x \mapsto x^p$, avec $p \geq 1$, on montre facilement que la norme d'Orlicz coïncide avec la norme de l'espace $L^p(\Omega, \mathcal{A}, \mathbb{P})$;
3. S'il existe C_0 tel que $\mathbb{E}\phi\left(\frac{|Z|}{C_0}\right) < \infty$, alors on peut montrer que la norme d'Orlicz de Z est finie. En effet, il suffit de voir que, grâce à un argument de convergence dominée, $\mathbb{E}\phi\left(\frac{|Z|}{C}\right) \rightarrow \mathbb{E}\phi(0) < 1$, lorsque $C \rightarrow \infty$, et de conclure par croissance de ϕ (en déduisant l'existence d'une constante positive C_1 inférieure à C_0 pour laquelle $\mathbb{E}\phi(|Z|/C_1) \leq 1$).

La norme d'Orlicz permet ainsi d'obtenir des inégalités reliant les queues de distributions à la fonction ϕ considérée par l'utilisation du lemme qui suit.

Lemme A.1.1. Soit ϕ une fonction convexe croissante positive telle que $\phi(0) \in [0, 1[$ et Z une variable aléatoire de norme d'Orlicz finie. Alors, pour tout

réel positif t ,

$$\mathbb{P}(|Z| \geq t) \leq \frac{1}{\phi\left(\frac{t}{C}\right)},$$

avec $C = \|Z\|_\phi$.

Démonstration :

Pour tout t réel positif, on peut écrire

$$\begin{aligned} \frac{\phi\left(\frac{|Z|}{C}\right)}{\phi\left(\frac{t}{C}\right)} &= \mathbb{1}_{\{|Z| \geq t\}} \frac{\phi\left(\frac{|Z|}{C}\right)}{\phi\left(\frac{t}{C}\right)} + \mathbb{1}_{\{|Z| < t\}} \frac{\phi\left(\frac{|Z|}{C}\right)}{\phi\left(\frac{t}{C}\right)} \\ &\geq \mathbb{1}_{\{|Z| \geq t\}}. \end{aligned}$$

On en déduit que $\mathbb{P}(|Z| \geq t) \leq \frac{1}{\phi\left(\frac{t}{C}\right)}$, car $\mathbb{E}\phi\left(\frac{|Z|}{C}\right) \leq 1$, par définition de la norme d'Orlicz. \square

À l'aide de fonctions convexes ϕ appropriées, il nous est alors possible de contrôler le comportement des queues de distributions par le biais d'inégalités exponentielles. Considérons ainsi l'application ψ définie comme suit :

$$\psi : \begin{cases} \mathbb{R}_+ & \rightarrow \mathbb{R} \\ x & \mapsto \frac{1}{5} \exp(x^2). \end{cases}$$

On obtient, par le lemme A.1.1, que pour tout t réel positif, et toute variable aléatoire Z fixée admettant une norme d'Orlicz finie,

$$\mathbb{P}(|Z| \geq t) \leq 5 \exp\left(-\frac{t^2}{\|Z\|_\psi^2}\right). \tag{A.1}$$

Il nous reste alors à adapter cette inégalité pour des maxima de suites de variables aléatoires, c'est dans cette optique que nous introduisons la notion de pseudo-dimension.

A.2 Pseudo-dimension

Considérons n variables aléatoires indépendantes ξ_1, \dots, ξ_n d'un ensemble Ξ abstrait, *i.e.* pour une σ -algèbre sur Ξ , chaque ξ_i est une application mesurable d'un espace probabilisé $(\Omega, \mathcal{A}, \mathbb{P})$ vers Ξ . Pour tout i , les $\{\xi_i\}$ définissent une mesure de probabilité sur Ξ . On définit la mesure empirique comme étant celle qui charge chaque point $\xi_1(\omega), \dots, \xi_n(\omega)$ par la masse $1/n$. Toute fonction f sur Ξ détermine alors une variable aléatoire : $\mathbb{E}_n f = n^{-1} \sum_{i=1}^n f(\xi_i(\omega))$. On cherche à généraliser les résultats de convergence lorsque la fonction f appartient à une classe de fonctions \mathcal{F} sur Ξ .

Soit la famille paramétrique $\mathcal{F} = \{f(\cdot, t); t \in T\}$. Si ξ_1, \dots, ξ_n sont n variables indépendantes, avec pour ω dans Ω et t dans T , $f_i(\omega, t) = f(\xi_i(\omega), t)$, nous nous intéressons plus particulièrement à

$$\mathcal{F}_n = \{(f_{1,n}(\cdot, t), \dots, f_{n,n}(\cdot, t)); t \in T\},$$

famille triangulaire de processus. La famille \mathcal{F} sera ici une famille paramétrique de dimension finie. De plus, dans nos applications, nous considérons essentiellement le cas où T est un intervalle réel borné. Le but est alors de construire des inégalités maximales pour

$$S_n(\cdot, t) - M_n(t) = \sum_{i=1}^n f_{n,i}(\cdot, t) - \mathbb{E} \sum_{i=1}^n f_{n,i}(\cdot, t),$$

ou pour

$$\Delta_n(\cdot) = \sup_{t \in T} |S_n(\cdot, t) - M_n(t)|.$$

Dans un premier temps, [Pol90] choisit d'introduire une nouvelle variable aléatoire et de se ramener à l'étude d'un sous-ensemble de \mathbb{R}^n : si l'on fixe ω dans Ω et l'on pose $\mathcal{F}_{n,\omega} = \{(f_{n1}(\omega, t), \dots, f_{n,n}(\omega, t)); t \in T\}$, $\mathcal{F}_{n,\omega}$ est alors un sous-ensemble de \mathbb{R}^n , espace par exemple muni des normes ℓ^1 ou de ℓ^2 , notées respectivement $|\cdot|_1$ et $|\cdot|_2$. Pour $\sigma_1, \dots, \sigma_n$, des variables aléatoires indépendantes pouvant prendre les valeurs 1 et -1 avec probabilité $1/2$, il introduit alors, sur un espace probabilité enrichi, un processus plus général :

$$\begin{aligned} L_n(\sigma, \omega) &= \sup_{t \in T} \left| \sum_{i=1}^n \sigma_i f_i(\omega, t) \right| \\ &= \sup_{f \in \mathcal{F}_{n,\omega}} |\sigma \cdot f|, \end{aligned}$$

où $\sigma = (\sigma_1, \dots, \sigma_n) \in \{-1; 1\}^n$, et $\sigma \cdot f$ désigne le produit scalaire usuel dans \mathbb{R}^n .

Exemple (partie 1/3)

Au cours de ce paragraphe introductif aux inégalités maximales, nous illustrerons les différentes propositions par le biais de l'exemple simple du processus empirique S_n , défini par

$$S_n(\omega, t) = \frac{1}{n} \sum_{i=1}^n \mathbb{1}_{\{X_i(\omega) \leq t\}},$$

où $(X_n)_{n \geq 1}$ est une suite de variables aléatoires indépendantes réelles et $t \in T$. Avec les notations précédentes, on aura donc :

$$\begin{aligned} f_{n,i}(\omega, t) &= \frac{1}{n} \mathbb{1}_{\{X_i(\omega) \leq t\}}, \text{ pour } t \in T \text{ et } \omega \in \Omega, \\ \mathcal{F}_n &= \{(f_{n,1}(\cdot, t), \dots, f_{n,n}(\cdot, t)); t \in T\}, \\ \mathcal{F}_{n,\omega} &= \{(f_{n,1}(\omega, t), \dots, f_{n,n}(\omega, t)); t \in T\}, \text{ pour } \omega \in \Omega. \end{aligned} \quad (\text{A.2})$$

Dans ce cas, $M_n(t) = \frac{1}{n} \sum_{i=1}^n F_{X_i}(t)$, où F_{X_i} représente la fonction de répartition de la variable aléatoire X_i , pour i un entier compris entre 1 et n . Notre but est alors de montrer la convergence complète vers 0 de $\sup_{t \in T} |S_n(\cdot, t) - M_n(t)|$, lorsque n tend vers l'infini.

Définition A.2.1. $D(\varepsilon, T_0)$ pour un sous-ensemble T_0 d'un espace métrique (T, d) est défini comme étant le plus grand entier m pour lequel il existe des points t_1, \dots, t_m de T_0 tels que $d(t_i, t_j) > \varepsilon$ pour $i \neq j$.

Ce sont les deux résultats fondamentaux suivants qui permettent d'établir des inégalités maximales.

Théorème A.2.1. Pour tout sous-ensemble \mathcal{F} de \mathbb{R}^n ,

$$\left\| \sup_{f \in \mathcal{F}} |\sigma \cdot f| \right\|_{\psi} \leq 9 \int_0^{\delta} \sqrt{\log D(x, \mathcal{F})} dx, \quad (\text{A.3})$$

où $\delta = \sup_{f \in \mathcal{F}} |f|_2$.

Démonstration : voir [Pol90, chapitre 3, p.11].

Théorème A.2.2. Pour toute fonction convexe croissante ϕ ,

$$\mathbb{E} \phi \left(\sup_{t \in T} |S_n(\cdot, t) - M_n(t)| \right) \leq \mathbb{E} \phi \left(2 \sup_{f \in \mathcal{F}_{n,\omega}} |\sigma \cdot f| \right). \quad (\text{A.4})$$

Démonstration : voir [Pol90, chapitre 2, p.7].

Définition A.2.2. Soit $x \in \mathbb{R}^k$ et J un sous-ensemble de $I = \{1, \dots, k\}$. On définit la J -ème orthante de x comme l'ensemble des points y de \mathbb{R}^k pour lesquels :

$$\begin{aligned} y_i &> x_i & \text{si } i \in J, \\ y_i &< x_i & \text{si } i \in I \setminus J. \end{aligned}$$

On dira d'un sous-ensemble de \mathbb{R}^k qu'il occupe la J -ème orthante de x s'il contient au moins un point dans cette orthante. Un sous-ensemble de \mathbb{R}^k entoure x s'il occupe les 2^k orthantes définies par x .

Définition A.2.3. Soit $I = (i(1), \dots, i(k))$ un k -uplets d'entiers compris entre 1 et n . On appelle I -projection du point (y_1, \dots, y_n) de \mathbb{R}^n , le point $(y_{i(1)}, \dots, y_{i(k)})$, même si tous les entiers $i(1), \dots, i(k)$ ne sont pas distincts.

Si, de plus, les $i(1), \dots, i(k)$ sont distincts, on dira que la projection est propre.

Définition A.2.4. \mathcal{F} sous-ensemble de \mathbb{R}^n admet une pseudo-dimension d'au plus V si pour chaque point x de \mathbb{R}^{V+1} , aucune projection propre de \mathcal{F} n'entoure x .

Exemple (partie 2/3)

Pour illustrer ce concept de pseudo-dimension, nous allons montrer que $\mathcal{F}_{n,\omega}$ (voir la définition donnée p.117) a une pseudo-dimension au plus égale à 1. Soit $x = (x_1, x_2) \in \mathbb{R}^2$.

On souhaite montrer qu'aucune projection propre de $\mathcal{F}_{n,\omega}$ n'entoure le point x du plan. Ici $I = (i(1), i(2))$ où $i = i(1)$ et $j = i(2)$ sont des entiers distincts compris entre 1 et n . Une I -projection propre d'un point $(f_1(t), \dots, f_n(t))$ de \mathcal{F} est donnée par le point $(f_i(t), f_j(t))$, donc les projections propres sont les $\{(f_i(t), f_j(t)); t \in T\}$, avec i et j distincts.

Considérons une projection propre $\mathcal{P} = \{(f_i(t), f_j(t)); t \in T\}$ quelconque de \mathcal{F} . Par l'absurde, supposons qu'elle entoure le point x du plan, *i.e.* qu'elle occupe les 4 orthantes définies par x , donc contient au moins un point dans chaque orthante. L'ensemble J de la définition A.2.2 peut prendre les valeurs suivantes : $\emptyset, \{1\}, \{2\}, \{1; 2\}$.

À chaque valeur possible de l'ensemble J correspond un point de \mathcal{P} qui est dans la J -ème orthante de x . Notons $(f_i(t_1), f_j(t_1)), \dots, (f_i(t_4), f_j(t_4))$ les points dans chacune des quatre orthantes de x .

En particulier, on obtient

$$\begin{aligned} f_i(t_2) &> x_1 & \text{et } f_j(t_2) &< x_2, \\ f_i(t_3) &< x_1 & \text{et } f_j(t_3) &> x_2. \end{aligned}$$

D'où on obtient $t_2 > t_3$ et $t_3 > t_2$ par croissance des f_j , on aboutit alors à une contradiction. \square

La démonstration effectuée ici dans le cas d'indicatrices croissantes peut en fait être généralisée pour les fonctions croissantes (c'est en effet l'unique propriété des f_i qui est utilisée pour calculer la pseudo-dimension dans l'exemple ci-dessus), ainsi nous pouvons par exemple reprendre [BGY97, lemme A.2, p.679] :

Lemme A.2.1. *Soit T , un sous-ensemble de la droite réelle. On suppose que, pour i entier, f_i désigne une fonction croissante sur T . Alors la famille $\{(f_1(t), \dots, f_n(t)); t \in T\}$ est de pseudo-dimension au plus 1.*

Nous allons désormais montrer comment les relations A.3 et A.4, combinées avec des propriétés géométriques des sous-ensembles de \mathbb{R}^n , conduisent à des inégalités maximales.

A.3 Inégalités maximales

Introduisons tout d'abord quelques notations et définitions :

Définition A.3.1.

1. Pour x et y deux vecteurs de \mathbb{R}^n , $x \odot y$ désigne le produit terme à terme de x et y , de plus, si $\mathcal{F} \subset \mathbb{R}^n$, $x \odot \mathcal{F} = \{x \odot y; y \in \mathcal{F}\}$.
2. Si \mathcal{F} est un sous-ensemble de \mathbb{R}^n , on appelle enveloppe de \mathcal{F} tout vecteur $\mathbb{F} = (F_1, \dots, F_n)$ tel que, pour tout élément $f = (f_1, \dots, f_n)$ de \mathcal{F} , on ait $|f_i| \leq F_i$, pour i entier de $\llbracket 1, n \rrbracket$.

Définition A.3.2.

Soit un tableau triangulaire de processus $\{f_{ni}(\omega, t); t \in T, 1 \leq i \leq k_n\}_{n \geq 1}$. On dit qu'il est contrôlable par rapport à l'enveloppe $\mathbb{F}_n(\omega)$, s'il existe une fonction déterministe λ , appelée borne de capacité, telle que :

- (i) $\int_0^1 \sqrt{\log \lambda(x)} dx < \infty$,
- (ii) $D(x|\alpha \odot \mathbb{F}_n(\omega)|_2, \alpha \odot \mathcal{F}_{n\omega}) \leq \lambda(x)$, pour $0 < x \leq 1$, tout ω , tout vecteurs α de constantes positives, et tout entier naturel n .

Dans le cas particulier où la borne de capacité a pour forme $\lambda(x) = A(1/x)^W$, avec A et W des constantes, on dira que les processus sont euclidiens.

Théorème A.3.1. Une suite de processus de pseudo-dimension finie est euclidienne et donc contrôlable. Si \mathcal{F} est un sous-ensemble borné de \mathbb{R}^n d'enveloppe \mathbb{F} et de pseudo-dimension au plus V , alors, il existe des constantes $A > 0$ et $W > 0$, dépendant uniquement de V , telles que

$$D(x|\alpha \odot \mathbb{F}|_2, \alpha \odot \mathcal{F}) \leq A \left(\frac{1}{x}\right)^W,$$

pour $x \in]0, 1]$ et tout vecteur α de constantes positives.

Démonstration : voir [Pol90, corollaire p.20]. □

Proposition A.3.1. Supposons que l'on ait :

- (i) $\delta_{n\omega} = \sup_{f \in \mathcal{F}_{n,\omega}} |f|_2 = \mathcal{O}\left(\frac{1}{\sqrt{n}}\right)$, uniformément en $\omega \in \Omega$,
- (ii) $\mathcal{F}_{n,\omega}$ est euclidienne,

alors, il existe des constantes c_1 et c_2 telles que

1. $\mathbb{P}(\Delta_n \geq \varepsilon) \leq c_1 \exp(-c_2 n)$,
2. $\Delta_n \xrightarrow{c} 0$, lorsque $n \rightarrow \infty$.

Démonstration :

Par hypothèse, on sait qu'il existe une constante positive C telle que $\delta_{n\omega} \leq Cn^{-1/2}$. Par ailleurs, comme $\mathcal{F}_{n,\omega}$ est euclidienne, il existe des constantes positives A et W telles que :

$$\begin{cases} D(x|\alpha \odot \mathbb{F}_n(\omega)|_2, \alpha \odot \mathcal{F}_{n\omega}) \leq \lambda(x), & \forall x, \\ \Lambda(1) = \int_0^1 \sqrt{\log \lambda(x)} dx < \infty, \end{cases}$$

avec $\lambda(x) = A(1/x)^W$.

Soit $J_n(\omega) = 9 \int_0^{\delta_{n\omega}} \sqrt{\log D(x, \mathcal{F}_{n,\omega})} dx$. Si on effectue le changement de variable $x = u|\mathbb{F}_n(\omega)|_2$, on obtient :

$$\begin{aligned} J_n(\omega) &= 9|\mathbb{F}_n(\omega)|_2 \int_0^{\frac{\delta_n}{|\mathbb{F}_n(\omega)|_2}} \sqrt{\log D(u|\mathbb{F}_n(\omega)|_2, \mathcal{F}_{n,\omega})} du \\ &\leq 9|\mathbb{F}_n(\omega)|_2 \int_0^1 \sqrt{\log D(u|\alpha \odot \mathbb{F}_n(\omega)|_2, \alpha \odot \mathcal{F}_{n,\omega})} du \\ &\leq 9 \frac{C}{\sqrt{n}} \Lambda(1) \\ &\leq K_n = C'n^{-1/2}, \end{aligned}$$

où, dans la seconde ligne, α est un vecteur dont chacune des coordonnées vaut 1.

Rappelons que

$$L_n(\omega, \sigma) = \sup_{f \in \mathcal{F}_{n,\omega}} |\sigma \cdot f|,$$

et que, par le théorème A.2.1 et le calcul qui précède, on a

$$\|L_n\|_\psi \leq K_n. \tag{A.5}$$

Par le théorème A.2.2, nous savons d'autre part que :

$$\mathbb{E}(\phi(\Delta_n)) \leq \mathbb{E}(\phi(2L_n)).$$

D'où on déduit par (A.5) que :

$$\|\Delta_n\|_\psi \leq 2K_n.$$

Vient alors, pour $\varepsilon > 0$, en utilisant (A.1) :

$$\mathbb{P}(|\Delta_n| > \varepsilon) \leq 5 \exp\left(\frac{-n\varepsilon^2}{4C'^2}\right),$$

qui conduit au résultat 1. Le résultat 2. provient du fait que le terme de droite de l'inégalité ci-dessus est celui d'une série convergente, ce qui démontre cette proposition. \square

Exemple (partie 3/3)

Revenons à l'exemple du paragraphe précédent. On sait déjà que la sous-famille

considérée est de pseudo-dimension 1, donc qu'elle est euclidienne, par la remarque A.3.1. Il nous reste alors à établir la condition (i) pour vérifier les conditions de la proposition A.3.1. Soit $f \in \mathcal{F}$. Alors

$$f = (f_{n,1}(\cdot, t), \dots, f_{n,n}(\cdot, t)),$$

et

$$|f|_2^2 = \sum_{i=1}^n \frac{1}{n^2} \mathbb{1}_{\{x_i \leq t\}} \leq \frac{1}{n}.$$

D'où $\delta_n \leq \frac{1}{\sqrt{n}}$. Finalement, on obtient la convergence complète souhaitée en appliquant la proposition A.3.1.

Nous aurons besoin de certains résultats de stabilité pour les familles euclidiennes. Ils sont répertoriés dans les lemmes qui suivent. Nous utilisons des résultats concernant les familles euclidiennes de processus, et non contrôlables contrairement à la version originale de ces lemmes, que l'on peut trouver dans [BGY97]. Pour les démontrer, on se reporte à [Pol90, Chapitre 5], et l'on remarque que les sous-ensembles de pseudo-dimension finie sont euclidiens.

Lemme A.3.1. *Si \mathcal{F} et \mathcal{G} sont deux sous-ensembles euclidiens de \mathbb{R}^n par rapport à une enveloppe commune \mathbb{F} , alors la somme terme à terme $\mathcal{F} \oplus \mathcal{G}$ est euclidienne par rapport à l'enveloppe $2\mathbb{F}$.*

Démonstration : voir [Pol90, Lemma 4.9 p.20]. □

Lemme A.3.2. *Soit T un sous-ensemble de \mathbb{R} . On suppose que $f_i(\cdot)$ est monotone pour tout $i = 1, \dots, n$. Alors $\mathcal{F} = \{(f_1(t), \dots, f_n(t)); t \in T\}$ a une pseudo-dimension au plus 1, et est donc euclidien.*

Démonstration : voir [BGY97, p.97]. □

Lemme A.3.3. *Si \mathcal{F} et \mathcal{G} sont deux sous-ensembles euclidiens de \mathbb{R}^n , alors leur produit terme à terme $\mathcal{F} \odot \mathcal{G} = \{f \odot g; f \in \mathcal{F}, g \in \mathcal{G}\}$ l'est aussi.*

Démonstration : voir [Pol90, Lemma 5.3]. □

Annexe B

Résultats et inégalités diverses

ette partie très succincte a pour but dans un premier temps de donner un corollaire au théorème issu de [SW86], puis de rappeler le résultat figurant dans l'article [DG00] utilisé dans la démonstration du paragraphe 5.2 dans le cadre multidimensionnel.

B.1 Inégalité portant sur les martingales

Pour M , une martingale locale sur $[0, \tau]$ de valeur nulle en 0, on définit :

- $\|M\|_0^\tau = \sup_{t \in [0, \tau]} |M(t)|$,
- pour $t \in [0, \tau]$, $\Delta M(t) = M(t) - M(t^-)$.

Théorème B.1.1. *Soit M une martingale locale, localement uniformément intégrable et nulle au point 0, avec $\|\Delta M\|_0^\tau \leq c$. Alors pour λ et θ réels strictement positifs, on a*

$$\mathbb{P}(\|M\|_0^\tau \geq \lambda) \leq \exp\left(-\frac{\lambda^2}{2\theta}\psi\left(\frac{\lambda c}{\theta}\right)\right) + \mathbb{P}(\langle M \rangle(\tau) \geq \theta),$$

$$\text{où : } \begin{cases} \psi : x \mapsto \frac{2h(1+x)}{x^2}, \\ h : x \mapsto x(\log x - 1) + 1. \end{cases}$$

Démonstration : voir [SW86, p.900]. □

Corollaire B.1.1. *Soit $(M_n)_{n \geq 1}$, une suite de martingales locales, localement uniformément intégrables et nulles en 0. Supposons que l'on ait, pour $n \geq 1$,*

$$\|\Delta M_n\|_0^\tau \leq \frac{c}{\sqrt{n}},$$

où c est une constante positive. Si

$$\sum_{n \geq 1} \mathbb{P}(\langle M_n \rangle(\tau) \geq n^{-1/4})$$

est une somme finie, alors $(\|M_n(t)\|_0^\tau)_{n \geq 1}$ converge complètement vers 0.

Démonstration : Considérons $(M_n)_{n \geq 1}$, une suite de martingales locales, localement uniformément intégrables et nulles en 0. On obtient dans ce cas, pour $n \geq 1$, avec le théorème B.1.1 et $\theta = \theta_n = n^{-1/4}$:

$$\mathbb{P}(\|M_n\|_0^\tau \geq \lambda) \leq \exp\left(-\frac{n^{1/4}\lambda^2}{2} \psi\left(\frac{\lambda c}{n^{1/4}}\right)\right) + \mathbb{P}(\langle M_n \rangle(\tau) \geq n^{-1/4}).$$

On remarque que la fonction ψ a pour limite 1 en 0. On en déduit donc que la série de terme général

$$\exp\left(-\frac{n^{1/4}\lambda^2}{2} \psi\left(\frac{\lambda c}{n^{1/4}}\right)\right),$$

converge. L'hypothèse sur $\mathbb{P}(\langle M_n \rangle(\tau) \geq n^{-1/4})$ permet alors de conclure. \square

B.2 Rappels de résultats

Lemme B.2.1. Soit $(A_n)_{n \geq 1}$ et $(B_n)_{n \geq 1}$, deux suites de matrices définies positives de $\mathcal{M}_p(\mathbb{R})$, et $(a_n)_{n \geq 1}$ et $(b_n)_{n \geq 1}$, les suites de valeurs propres minimales correspondantes. Alors, on a, pour $n \geq 1$,

$$\left|\frac{a_n}{b_n} - 1\right| \leq \frac{\|A_n - B_n\|}{b_n}, \quad (\text{B.1})$$

où $\|\cdot\|$ désigne une norme de $\mathcal{M}_p(\mathbb{R})$.

Démonstration : voir [DG00, lemme A.2]. \square

Remarque B.2.1. Il est possible de montrer (B.1) si $(A_n)_{n \geq 1}$ est une suite de matrices semi-définies positives.

Annexe C

Condition d'Anscombe

On s'intéresse à la normalité asymptotique d'une suite de variables aléatoires donnée sous la forme Y_{T_n} , où $(T_n)_{n \geq 1}$ est une suite de temps d'arrêt, alors que l'on dispose d'un résultat de convergence sur $(Y_n)_{n \geq 1}$. C'est à cette fin qu'Anscombe introduisit la notion de continuité uniforme en probabilité, qui suffit à obtenir le résultat souhaité.

C.1 Continuité uniforme en probabilité et première condition d'Anscombe

On trouve dans [Ans52] un premier critère utilisant la notion de convergence uniforme en probabilité.

Définition C.1.1. $(Y_n)_{n \geq 1}$ est une suite de variables aléatoires uniformément continue en probabilité si et seulement si elle vérifie la condition suivante.

Il existe une suite de réels positifs $(w_n)_{n \geq 1}$ telle que, pour tout $\varepsilon > 0, \eta > 0$, il existe ν et $c \geq 0$ tels que

$$\forall n \geq \nu, \mathbb{P}(|Y_n - Y_{n'}| < \varepsilon w_n, \forall n'; |n' - n| < cn) > 1 - \eta,$$

ce qui est équivalent à

$$\forall n \geq \nu, \mathbb{P}\left(\max_{\{n'; |n'-n| < cn\}} |Y_n - Y_{n'}| < \varepsilon w_n\right) > 1 - \eta.$$

Il démontre alors un premier théorème.

Théorème C.1.1. Soit $(Y_n)_{n \geq 1}$ une suite de variables aléatoires. On suppose qu'il existe un réel θ , une suite de nombres positifs $(w_n)_{n \geq 1}$ et une fonction de répartition F tels que :

(C1) (Convergence en distribution de $(Y_n)_{n \geq 1}$)
 Pour tout point de continuité x de F ,

$$\mathbb{P}(Y_n - \theta \leq xw_n) \rightarrow F(x), \text{ quand } n \rightarrow \infty,$$

(C2) $(Y_n)_{n \geq 1}$ est uniformément continue en probabilité.

Soit $(n_r)_{r \geq 1}$, une suite croissante d'entiers positifs qui tend vers l'infini, et soit $(N_r)_{r \geq 1}$, une suite de variables aléatoires à valeurs entières positives telles que

$$\frac{N_r}{n_r} \xrightarrow{P} 1, \text{ quand } r \rightarrow \infty.$$

Alors $(Y_n)_{n \geq 1}$ est telle que

$$\mathbb{P}(Y_{N_r} - \theta \leq xw_{n_r}) \rightarrow F(x), \text{ lorsque } r \rightarrow \infty,$$

pour tout point de continuité x de F .

On remarque ici que la vitesse de convergence apparaît dans la suite $(w_n)_{n \geq 1}$.

C.2 Papiers postérieurs

Deux années plus tard, Anscombe [Ans54] publie un article plus adapté à notre situation et donc plus adapté aux méthodes séquentielles d'arrêt. Sa contribution sera surtout à l'origine d'un article où la normalité asymptotique est cette fois démontrée dans le cas particulier d'intervalles de confiance séquentiels relatifs à l'estimation de la moyenne d'une loi quelconque [CR65]. La notion d'efficacité asymptotique de la règle d'arrêt y est introduite, et la normalité asymptotique établie à l'aide des résultats d'Anscombe. Cet article sera ensuite à l'origine de nombreuses études sur les règles d'arrêt. Ainsi, Gleser s'intéresse plus particulièrement aux cas des régressions linéaires et utilise une fois encore la condition d'Anscombe [Gle65].

Battacharya et Mallik [BM73] choisissent de démontrer la normalité asymptotique de différentes procédures d'arrêt, à l'aide des résultats de Chow et Robbins [CR65].

On voit ensuite apparaître plusieurs types de règles d'arrêt. La continuité uniforme en probabilité est souvent nécessaire dans les démonstrations (voir par exemple [Cha96, Cha01]), même si certains résultats peuvent être établis à l'aide de techniques plus adaptées aux différents modèles, comme le principe d'invariance faible de la suite $(T_n)_{n \geq 1}$ [GMS97, p.318].

Nous nous sommes intéressés dans le cadre de cette thèse plus particulièrement aux techniques proposées dans [DG00] et dans [GMS97], où l'on peut trouver des définitions de règles d'arrêts aussi bien dans le cas paramétrique que non paramétrique. Nous définirons ainsi la régularité.

Définition C.2.1. Soit $(T_n)_{n \geq 1}$, une suite de variables aléatoires entières et positives. On suppose qu'il existe une suite déterministe $(t_n)_{n \geq 1}$ d'entiers positifs qui croît vers l'infini telle que

$$\frac{T_n}{t_n} \xrightarrow{P} 1, \text{ lorsque } n \rightarrow \infty.$$

$(T_n)_{n \geq 1}$ est alors appelée suite régulière de variables aléatoires.

Nous utiliserons la continuité uniforme sous la forme suivante.

Théorème C.2.1. Soit $(Y_n)_{n \geq 1}$, une suite de variables aléatoires telle que :

$$Y_n \xrightarrow{D} \mathcal{N}(0, 1), \text{ lorsque } n \rightarrow \infty,$$

et telle que pour $\varepsilon > 0$, il existe $\delta > 0$ tel que, pour tout $n \geq 1$, on ait :

$$\mathbb{P} \left(\max_{-\delta n \leq k \leq n\delta} |Y_{n+k} - Y_n| > \varepsilon \right) < \varepsilon. \quad (\text{C.1})$$

On a alors :

1. $Y_{T_n} - Y_{t_n} \xrightarrow{P} 0$, lorsque $n \rightarrow \infty$,
2. $Y_{T_n} \rightarrow \mathcal{N}(0, 1)$, lorsque $n \rightarrow \infty$.

Démonstration : Soit $\varepsilon > 0$ et $0 < \delta < 1$, donné par la condition d'Anscombe. Pour n assez grand, on écrit :

$$\mathbb{P} (|Y_{T_n} - Y_{t_n}| > \varepsilon) \leq \mathbb{P} \left(\left| \frac{T_n}{t_n} - 1 \right| > \delta \right) + \mathbb{P} \left(\max_{\{k; |k-t_n| \leq \delta t_n\}} |Y_{k+t_n} - Y_{t_n}| > \varepsilon \right).$$

Le premier membre de droite converge vers 0, par régularité de la suite $(T_n)_{n \geq 1}$. Pour montrer la convergence du second membre, on utilise la condition d'Anscombe. Finalement, le second terme de l'inégalité précédente converge vers 0, puisque $(t_n)_{n \geq 1}$ est une suite déterministe divergeant vers l'infini. Le second point se déduit immédiatement du premier. \square

Remarque C.2.1. Il suffit de vérifier la propriété (C.1) à partir d'un certain rang pour établir la condition d'Anscombe [GMS97]. En outre, il n'est pas nécessaire de vérifier cette propriété pour les valeurs négatives de k (voir [GMS97, p.318]).

Deuxième partie
Tests d'homogénéité

Chapitre 8

Cadre de l'étude

Sommaire

8.1	Modèle adopté et propriétés	133
8.1.1	Définitions et notations	133
8.1.2	Estimations et propriétés	134
8.2	Test d'homogénéité	135
8.2.1	Hypothèses	136
8.2.2	Calcul de la statistique de test	136

onsidérons un modèle non paramétrique caractérisé uniquement par sa fonction de risque α . On suppose en outre que l'on dispose de variables explicatives Z que l'on introduit au sein de la fonction α , qui est alors une fonction du temps et des variables explicatives, elles-même éventuellement dépendante du temps (voir [MU90]). À travers la variable Z , ce modèle permet de tenir compte d'une certaine hétérogénéité entre les individus. Cependant, il ne permet pas de tenir compte d'un éventuel effet de groupe, qui serait alors lié à un ensemble d'individus, non observable en pratique. Intervient ainsi un modèle plus général, permettant de prendre en compte cet effet et obtenu en ajoutant un facteur multiplicatif à la fonction de risque. Dans le cas d'un événement d'intérêt relatif à une panne, cette variable peut être vue comme une fragilité qui augmente le risque de panne. Notons que les variables d'effet ne sont pas nécessairement indépendantes, puisque deux individus du même groupe peuvent bénéficier du même effet. Adopter un modèle à effets aléatoires lorsque ce n'est pas nécessaire induit une diminution de l'efficacité pour l'estimation des paramètres ainsi qu'une plus grande complexité de formulation du modèle [NC99], c'est pourquoi il se révèle utile de tester la nécessité des variables d'effet. Pour ce faire, on fait intervenir une fonction score, qui sera notre statistique de test.

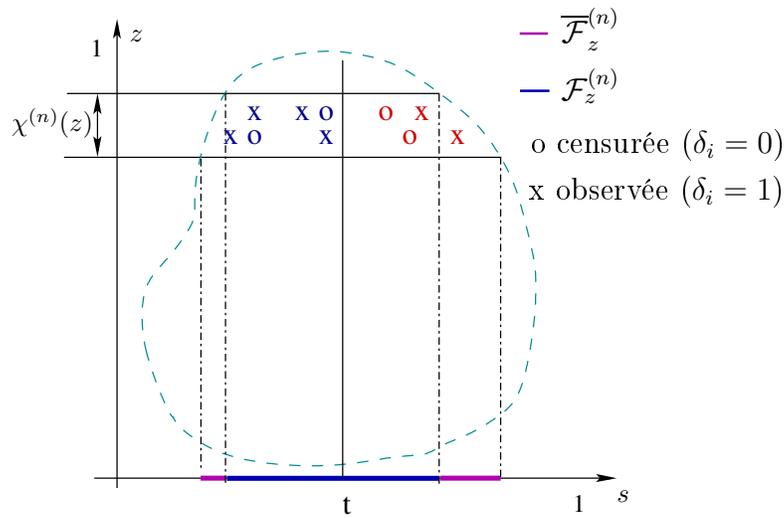


FIG. 8.1 – Description de $\mathcal{F}_z^{(n)}$ et $\overline{\mathcal{F}}_z^{(n)}$

8.1 Modèle adopté et propriétés

ette première partie vise à établir sa formulation, nous calculerons sa loi limite par la suite.

8.1.1 Définitions et notations

Considérons une variable explicative Z , à valeurs dans l'intervalle $[0, 1]$, ainsi qu'une variable aléatoire T de durée de vie, à valeurs dans $[0, 1]$. Supposons que conditionnellement à la donnée de la variable explicative $Z = z \in [0, 1]$, la durée de vie T admette pour fonction de risque $\alpha(\cdot; z)$. Soit $A(\cdot; z)$ la fonction de risque cumulé conditionnelle à z , définie pour $t \in [0, 1]$ par :

$$A(t; z) = \int_0^t \alpha(u; z) du.$$

On suppose en outre que α est une fonction lipchitzienne sur $[0, 1]^2$, de support $\text{supp}(\alpha)$ inclus dans $[0, 1]^2$.

Pour $z \in [0, 1]$ et $\{\chi_i^{(n)}; 1 \leq i \leq k_n\}$ une partition de $[0, 1]$, on note $\chi^{(n)}(z) = \chi_i^{(n)}$, pour i tel que $z \in \chi_i^{(n)}$. On adopte les notations suivantes :

$$\begin{aligned} \mathcal{F}_z^{(n)} &= \{s \in [0, 1]; \chi^{(n)}(z) \subset \text{supp}(\alpha(s, \cdot))\}, \\ \overline{\mathcal{F}}_z^{(n)} &= \{s \in [0, 1]; \exists u, v \in \chi^{(n)}(z) \text{ avec } \alpha(s, u) > 0 \text{ et } \alpha(s, v) = 0\}, \end{aligned}$$

ces deux ensembles sont représentés sur la figure 8.1. On note ω_n l'amplitude commune des intervalles de la partition choisie pour n observations.

Les données sont supposées censurées à droite et l'on note C , la variable aléatoire modélisant la censure. On observe alors les n premiers éléments d'une suite $(X_i, \delta_i, Z_i)_{i \geq 1}$ de représentants de (X, δ, Z) , où :

- $X = \bar{T} \wedge C$,
- $\delta = \mathbb{1}_{\{T \leq C\}}$,
- i est l'indice du i^e individu.

8.1.2 Estimations et propriétés

L'estimateur que nous présentons ici fut introduit par Beran [Ber81] dans le cadre de covariables ne dépendant pas du temps. Cet estimateur fut ensuite généralisé par Mac Keague et Utikal [MU90] au cas de covariables dépendantes du temps. Il peut être vu comme un estimateur de Nelson-Aalen, défini par tranches. Ces dernières années, de nombreux modèles et méthodes d'estimation sont venus compléter et étendre ces travaux. Citons entre autres [Dab97, GW91]. Nous allons maintenant rappeler l'un des principaux résultats établis dans [MU90].

Ainsi, on définit la martingale M_i , pour $1 \leq i \leq n$, pour $t \in [0, 1]$, par :

$$M_i(t) = M_i(t; z_i) = N_i(t) - \Lambda_i(t; z_i),$$

où

$$\begin{aligned} N_i(t) &= \mathbb{1}_{\{X_i \leq t, \delta_i = 1\}}, \\ \Lambda_i(t; z_i) &= \int_0^t Y_i(s) \alpha(s; z_i) ds, \\ Y_i(s) &= \mathbb{1}_{\{X_i \geq s\}}. \end{aligned}$$

Remarque 8.1.1. *Afin d'alléger les notations, on écrira parfois $B_i(t)$ en lieu et place de $B_i(t; z_i)$, pour B un processus quelconque. En outre, $B_i(t; z_i)$ désigne la quantité :*

$$B_i(t; Z) \Big|_{Z=z_i}.$$

On définit alors des comptages « par tranches » donnés par :

$$\begin{aligned} N^{(n)}(t; z) &= \sum_{j=1}^n \mathbb{1}_{\{z_j \in \mathcal{X}^{(n)}(z)\}} N_j(t), \\ Y^{(n)}(t; z) &= \sum_{j=1}^n \mathbb{1}_{\{z_j \in \mathcal{X}^{(n)}(z)\}} Y_j(t). \end{aligned}$$

On définit $\hat{A}^{(n)}$, l'estimateur de Beran [Ber81] de A de la façon suivante :

$$\hat{A}^{(n)}(t; z_i) = \int_0^t \frac{N^{(n)}(ds; z_i)}{Y^{(n)}(s; z_i)}.$$

Alors, pour

$$\hat{\Lambda}^{(n)}(t; z_i) = \int_0^t Y_i(s) \hat{A}^{(n)}(ds; z_i),$$

on définit le résidu de martingale

$$\hat{M}^{(n)}(t; z_i) = N_i(t) - \hat{\Lambda}^{(n)}(t; z_i).$$

Théorème 8.1.1. [MU90] Soit $z \in [0, 1]$. On suppose que les conditions suivantes sont vérifiées.

(A1) Il existe une fonction mesurable positive bornée $g(\cdot; z)$ définie sur le support de $\alpha(\cdot, z)$ telle que

$$\int_{\mathcal{F}_z^{(n)}} \left| \frac{n\omega_n}{Y^{(n)}(s; z)} - g(s; z) \right| ds \xrightarrow{P} 0,$$

lorsque n tend vers l'infini.

(A2) $\text{Leb} \left\{ s \in \mathcal{F}_z^{(n)}; Y^{(n)}(s; z) = 0 \right\} = o_P(1/\sqrt{n\omega_n})$, lorsque $n \rightarrow +\infty$,

(A3) $\text{Leb} \left(\overline{\mathcal{F}}_z^{(n)} \right) = \mathcal{O}(\omega_n)$, lorsque $n \rightarrow +\infty$.

Si $n\omega_n \rightarrow \infty$ et $n\omega_n^2 \rightarrow 0$, on a :

$$\sqrt{n\omega_n} \left(\hat{A}(\cdot; z) - A(\cdot; z) \right) \xrightarrow{\mathcal{D}} U(\cdot, z),$$

dans $\mathcal{D}([0, 1])$, où $U(\cdot, z)$ est une martingale gaussienne continue, de moyenne nulle et de variance donnée par

$$\text{Var}(U(t; z)) = \int_0^t \alpha(s; z) g(s; z) ds, t \in [0, 1].$$

Remarque 8.1.2. La démonstration du théorème précédent donnée dans [MU90] requiert le résultat intermédiaire suivant, qu'on utilisera par la suite. Avec la notation :

$$\alpha^{(n)}(s; z) = \sum_{i=1}^n \mathbb{1}_{\{z_i \in \mathcal{X}^{(n)}(z)\}} Y_i(s) \alpha(s; z_i),$$

on a, uniformément en $z \in [0, 1]$ et $s \in \mathcal{F}_z^{(n)}$,

$$\alpha^{(n)}(s; z) = (\alpha(s; z) + \mathcal{O}_P(\omega_n)) Y^{(n)}(s; z).$$

8.2 Test d'homogénéité

près quelques hypothèses sur le modèle étudié, nous définissons la statistique de test à partir de la fonction score.

8.2.1 Hypothèses

On souhaite vérifier que les variables explicatives Z_i suffisent à expliquer la loi des durées de vie T_i . Pour tester cette hypothèse d'homogénéité, on suppose que la loi des données est perturbée par des effets aléatoires multiplicatifs non observables. Pour $s \in [0, 1]$, l'intervalle d'étude, on a donc :

$$\alpha(s; z_i, \varepsilon_i) = \exp(\varepsilon_i)\alpha(s; z_i),$$

où $\varepsilon_i = \sqrt{\theta}V_i$, avec θ un réel positif et V_i des variables aléatoires réelles satisfaisant les hypothèses données ci-dessous.

Le test consiste alors à tester $\mathcal{H}_0 : \ll \theta = 0 \gg$.

Hypothèses 5. En notant $\mathbf{V} = (V_1, \dots, V_n)^T$ de distribution G inconnue, on suppose que :

- $\mathbb{E}V_i = 0$, pour $1 \leq i \leq n$,
- $\mathbb{E}V_i^2 = 1$, pour $1 \leq i \leq n$,
- $\mathbb{E}(\mathbf{V}\mathbf{V}^T) = W$, où W est une matrice connue.

Remarque 8.2.1. Si l'on souhaite modéliser des effets de groupe, pour deux individus i et j du même groupe, on aura $w_{ij} = 1$, et w_{ij} sera nul pour deux individus de groupes distincts [CJG97]. Cela permet ainsi d'introduire directement une corrélation intra-groupe.

La première étape de ce travail consiste à calculer l'expression de la fonction de vraisemblance. À cette fin, on utilise des techniques de calcul similaires à celles utilisées dans [CJG97].

8.2.2 Calcul de la statistique de test

Nous venons de voir que le test d'homogénéité consiste à tester l'hypothèse :

$$\mathcal{H}_0 : \ll \theta = 0 \gg,$$

car sous cette hypothèse, il n'y a pas de « déviation » par rapport au modèle initial. Le test que nous proposons est basé sur l'idée très simple suivante. Si \mathcal{H}_0 est vraie, alors la log-vraisemblance des observations, vue comme une fonction de θ , devrait atteindre son maximum dans un voisinage de 0. Par conséquent, la dérivée en θ de la log-vraisemblance devrait, sous réserve de conditions de régularité, s'annuler dans un voisinage de 0. Le but est donc double, il faut :

1. estimer la dérivée de la log-vraisemblance en $\theta = 0$ (*i.e.* la fonction « score »), c'est l'objet de ce paragraphe,
2. établir la loi approchée de l'estimateur, de manière à rendre le test effectif. Ce sera l'objet du chapitre suivant.

Pour $T^{(n)}$, la fonction score pour n individus, il est possible de se ramener à l'écriture suivante :

$$T^{(n)} = \lim_{t \rightarrow \tau^-} T^{(n)}(t),$$

où $[0, \tau]$ est l'intervalle d'étude (comme nous l'avons vu au paragraphe précédent, $\tau = 1$ pour simplifier les notations).

Définissons la vraisemblance totale L_n observée qui est conditionnelle aux variables explicatives, et fonction du vecteur $\sqrt{\theta}\mathbf{v}$, où $\mathbf{v} = (v_1, \dots, v_n)$ est la valeur prise par le vecteur \mathbf{V} . Conditionnellement à la donnée des z_i et ε_i , les variables aléatoires de durée de vie sont indépendantes, ce qui nous permet de déduire que la fonction de vraisemblance s'écrit sous forme de produit, le i^e terme du produit modélisant la contribution de l'individu i à la vraisemblance. On peut alors écrire

$$L_n(\sqrt{\theta}\mathbf{v}) = \prod_{i=1}^n L_i(\sqrt{\theta}v_i) = \prod_{i=1}^n L_i(\varepsilon_i).$$

À l'aide de la formulation de la vraisemblance

$$L_n(\varepsilon_1, \dots, \varepsilon_n) = \prod_{i=1}^n L_i(\varepsilon_i),$$

on obtient, pour $1 \leq i \leq n$,

$$L_i(\varepsilon_i) = (e^{\varepsilon_i \alpha(X_i; z_i)})^{\delta_i} \exp\left(-\int_0^{X_i} e^{\varepsilon_i \alpha(u; z_i)} du\right).$$

Cela nous permet en outre de constater que la fonction $(\theta, v_i) \mapsto L_i(\sqrt{\theta}v_i)$, est deux fois dérivable sur $\mathbb{R}_+^* \times \mathbb{R}$.

Proposition 8.2.1. *Sous les hypothèses 5,*

$$T^{(n)}(t) = \sum_{i=1}^n \sum_{j=1}^n \int_0^t w_{ij} M_j(s-) dM_i(s) + \frac{1}{2} \sum_{i=1}^n \int_0^t dM_i(s).$$

Si W est la matrice identité, on a ainsi :

$$\begin{aligned} T^{(n)}(t) &= \frac{1}{2} \sum_{i=1}^n (M_i^2(t) - \Lambda(t; z_i)) \\ &= \sum_{i=1}^n \int_0^t (M_i(s-) + 1/2) M_i(ds). \end{aligned}$$

Démonstration :

Pour n entier supérieur à 1, on note $\boldsymbol{\varepsilon} = (\varepsilon_1, \dots, \varepsilon_n)$ et $\mathbf{v} = (v_1, \dots, v_n)$ (les v_i représenteront ici des variables muettes lors de formulations intégrales). La densité

connue est une densité conditionnelle à la donnée des ε_i , donc si l'on calcule la vraisemblance, on obtient :

$$\ell_n(\theta) = \int_{\mathbf{R}^n} L_n(\sqrt{\theta}\mathbf{v}) dG(\mathbf{v}).$$

Par définition de la statistique de score, on pose, conditionnellement à la donnée de z_1, \dots, z_n ,

$$T^{(n)} = \lim_{\theta \rightarrow 0} \frac{\partial \mathcal{L}_n}{\partial \theta}(\theta),$$

où $\mathcal{L}_n(\theta) = \log \ell_n(\theta)$. On obtient finalement

$$\frac{\partial \mathcal{L}_n}{\partial \theta}(\theta) = \frac{\frac{\partial}{\partial \theta} \int_{\mathbf{R}^n} L_n(\sqrt{\theta}\mathbf{v}) dG(\mathbf{v})}{\int_{\mathbf{R}^n} L_n(\sqrt{\theta}\mathbf{v}) dG(\mathbf{v})}.$$

On note alors :

$$\begin{aligned} \frac{\partial L_i}{\partial \theta}(\sqrt{\theta}v_i) &= \frac{v_i}{2\sqrt{\theta}} L'_i(\varepsilon_i), \text{ où } L'_i(\varepsilon_i) = \frac{dL_i}{d\varepsilon_i}(\varepsilon_i) = L'_i(\sqrt{\theta}v_i), \\ \frac{\partial L'_i}{\partial \theta}(\sqrt{\theta}v_i) &= \frac{v_i}{2\sqrt{\theta}} L''_i(\varepsilon_i), \text{ où } L''_i(\varepsilon_i) = \frac{d^2 L_i}{d\varepsilon_i^2}(\varepsilon_i). \end{aligned}$$

Il est alors possible d'appliquer le théorème de dérivation sous le signe somme, en remarquant simplement que dG est une mesure finie, que $\theta \mapsto L_i(\sqrt{\theta}v_i)$ est de classe C^1 sur \mathbb{R}_+^* et que sa dérivée par rapport à la variable θ est bornée pour θ élément d'un compact K de \mathbb{R}_+^* . On peut ainsi écrire

$$\begin{aligned} \frac{\partial}{\partial \theta} \int_{\mathbf{R}^n} L_n(\sqrt{\theta}\mathbf{v}) dG(\mathbf{v}) &= \int_{\mathbf{R}^n} \frac{\partial}{\partial \theta} \prod_{i=1}^n L_i(\varepsilon_i) dG(\mathbf{v}) \\ &= \int_{\mathbf{R}^n} \sum_{i=1}^n \frac{\partial L_i}{\partial \varepsilon_i}(\varepsilon_i) \prod_{j \neq i} L_j(\varepsilon_j) \frac{v_i}{2\sqrt{\theta}} dG(\mathbf{v}). \end{aligned}$$

Nous calculons désormais l'expression de $T^{(n)}$, conditionnelle à \mathbf{z} :

$$\begin{aligned} T^{(n)} &= \lim_{\theta \rightarrow 0} \frac{\partial \mathcal{L}_n}{\partial \theta}(\theta) \\ &= \frac{\lim_{\theta \rightarrow 0} \frac{\partial}{\partial \theta} \int_{\mathbf{R}^n} L_n(\sqrt{\theta}\mathbf{v}) dG(\mathbf{v})}{\lim_{\theta \rightarrow 0} \int_{\mathbf{R}^n} L_n(\sqrt{\theta}\mathbf{v}) dG(\mathbf{v})} \\ &= \frac{\frac{1}{2} \sum_{i=1}^n \int_{\mathbf{R}^n} \lim_{\theta \rightarrow 0} L'_i(\varepsilon_i) \prod_{j \neq i} L_j(\varepsilon_j) \frac{v_i}{\sqrt{\theta}} dG(\mathbf{v})}{L(0)} \\ &= \frac{1}{2} \sum_{i=1}^n \frac{1}{L(0)} \lim_{\theta \rightarrow 0} \left\{ \frac{1}{\sqrt{\theta}} \int_{\mathbf{R}^n} v_i L'_i(\varepsilon_i) \prod_{j \neq i} L_j(\varepsilon_j) dG(\mathbf{v}) \right\}. \end{aligned}$$

La limite du terme entre accolades ci-dessus est indéterminée, de la forme 0/0. En effet, le dénominateur n'est autre que la fonction racine et si l'on calcule la limite lorsque θ tend vers 0 du numérateur, on obtient un multiple de l'espérance de V_i (i entier compris entre 1 et n), supposée nulle. On souhaite ainsi utiliser la règle de l'Hôpital pour calculer la limite d'un quotient. Son utilisation est possible, puisque la fonction racine ainsi que le numérateur sont dérivables sur un voisinage ouvert à droite de 0, et que la dérivée de la fonction racine carrée est non nulle sur ce même voisinage. On procède désormais au calcul préalable de la dérivée du numérateur :

$$\begin{aligned}
& \frac{\partial}{\partial \theta} \sum_{i=1}^n \int_{\mathbf{R}^n} v_i L'_i(\varepsilon_i) \prod_{j \neq i} L_j(\varepsilon_j) dG(\mathbf{v}) \\
&= \sum_{i=1}^n \int_{\mathbf{R}^n} \frac{\partial}{\partial \theta} v_i L'_i(\varepsilon_i) \prod_{j \neq i} L_j(\varepsilon_j) dG(\mathbf{v}) \\
&= \sum_{i=1}^n \int_{\mathbf{R}^n} \frac{v_i^2}{2\sqrt{\theta}} L''_i(\varepsilon_i) \prod_{j \neq i} L_j(\varepsilon_j) dG(\mathbf{v}) \\
&+ \sum_{i=1}^n \int_{\mathbf{R}^n} \sum_{k \neq i} \frac{v_k}{2\sqrt{\theta}} v_i L'_i(\varepsilon_i) L'_k(\varepsilon_k) \prod_{j \neq i, k} L_j(\varepsilon_j) dG(\mathbf{v}).
\end{aligned}$$

En utilisant le fait que $L(0) = \prod_{i=1}^n L_i(0)$, on peut ensuite écrire que :

$$\begin{aligned}
T^{(n)} &= \frac{1}{2} \sum_{i=1}^n \left\{ \int_{\mathbf{R}^n} \frac{v_i^2 L''_i(0)}{L_i(0)} dG(\mathbf{v}) + \sum_{k \neq i} \int_{\mathbf{R}^n} \frac{v_k L'_k(0) v_i L'_i(0)}{L_k(0) L_i(0)} dG(\mathbf{v}) \right\} \\
&= \frac{1}{2} \sum_{i=1}^n \left\{ w_{ii} \frac{L''_i(0)}{L_i(0)} + \sum_{j \neq i} w_{ji} \frac{L'_j(0) L'_i(0)}{L_j(0) L_i(0)} \right\} \\
&= \frac{1}{2} \sum_{i=1}^n w_{ii} \frac{L''_i(0)}{L_i(0)} + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ji} \frac{L'_j(0) L'_i(0)}{L_i(0) L_j(0)} - \frac{1}{2} \sum_{i=1}^n w_{ii} \left(\frac{L'_i(0)}{L_i(0)} \right)^2.
\end{aligned}$$

Par ailleurs, on peut voir que

$$\frac{\frac{d^2 L_i}{d\varepsilon_i^2}(\varepsilon_i) \Big|_{\theta=0}}{L_i(0)} = \left(\frac{d \log L_i}{d\varepsilon_i}(\varepsilon_i) \Big|_{\theta=0} \right)^2 + \frac{d^2 \log L_i}{d\varepsilon_i^2}(\varepsilon_i) \Big|_{\theta=0}.$$

Il est alors possible de déduire que

$$T^{(n)} = \frac{1}{2} \sum_{i=1}^n w_{ii} \left(\frac{d^2 \log L_i}{d\varepsilon_i^2}(\varepsilon_i) \Big|_{\theta=0} \right) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ji} \frac{d \log L_j}{d\varepsilon_j}(\varepsilon_j) \Big|_{\theta=0} \frac{d \log L_i}{d\varepsilon_i}(\varepsilon_i) \Big|_{\theta=0}.$$

On peut exprimer les dérivées partielles de la manière suivante, pour $1 \leq i \leq n$:

$$\begin{aligned}
\frac{\partial \log L_i}{\partial \varepsilon_i}(\varepsilon_i) \Big|_{\theta=0} &= \lim_{t \rightarrow \infty} (\delta_i - \Lambda(t; z_i)) = N_i(\infty) - \Lambda(\infty; z_i), \\
\frac{\partial^2 \log L_i}{\partial \varepsilon_i^2}(\varepsilon_i) \Big|_{\theta=0} &= -\Lambda(\infty; z_i).
\end{aligned}$$

En injectant les expressions précédentes des dérivées des log-vraisemblances, on obtient :

$$T^{(n)} = \lim_{t \rightarrow \infty} T^{(n)}(t),$$

où

$$T^{(n)}(t) = \frac{1}{2} \sum_{i=1}^n -w_{ii} \Lambda(t; z_i) + \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n w_{ji} M_i(t) M_j(t).$$

On utilise ensuite la proposition D.0.1 pour avoir

$$\begin{aligned} M_i(t) M_j(t) &= \int_0^t M_i(s-) dM_j(s) + \int_0^t M_j(s-) dM_i(s) \quad (i \neq j), \\ M_i(t)^2 &= 2 \int_0^t M_i(s-) dM_i(s) + N_i(t). \end{aligned}$$

D'où :

$$\begin{aligned} T^{(n)}(t) &= -\frac{1}{2} \sum_{i=1}^n w_{ii} \Lambda(t; z_i) + \sum_{i=1}^n \sum_{j=1}^n \int_0^t \bar{w}_{ij} M_i(s-) dM_j(s) \\ &+ \sum_{i=1}^n \int_0^t w_{ii} M_i(s-) dM_i(s) + \frac{1}{2} \sum_{i=1}^n w_{ii} N_i(t) \\ &= \sum_{i=1}^n w_{ii} \int_0^t \left(M_i(s-) + \frac{1}{2} \right) dM_i(s) + \sum_{i=1}^n \sum_{j=1}^n \int_0^t \bar{w}_{ij} M_i(s-) dM_j(s) \\ &= \sum_{i=1}^n \sum_{j=1}^n \int_0^t w_{ij} M_j(s-) dM_i(s) + \frac{1}{2} \sum_{i=1}^n \int_0^t dM_i(s), \end{aligned}$$

avec $\bar{w}_{ij} = w_{ij} \mathbb{1}_{\{i \neq j\}}$. □

Nous disposons désormais d'une expression de la fonction score, notre démarche consiste alors à l'estimer par le biais de résidus de martingales et déterminer le comportement asymptotique de l'estimateur ainsi obtenu.

Pour estimer $T^{(n)}$, on remplace dans l'expression précédemment obtenue les martingales M_i par leurs résidus $\hat{M}_i^{(n)}$, ce qui nous permet d'obtenir pour

$$\hat{T}^{(n)} = \lim_{t \rightarrow \tau} \hat{T}^{(n)}(t),$$

où $[0, \tau]$ est l'intervalle d'étude,

$$\hat{T}^{(n)}(t) = \sum_{i=1}^n \sum_{j=1}^n \int_0^t w_{ij} \hat{M}^{(n)}(s-; z_j) \hat{M}^{(n)}(ds; z_i) + \frac{1}{2} \sum_{i=1}^n \int_0^t \hat{M}^{(n)}(ds; z_i).$$

On définit ainsi la statistique pour le test d'homogénéité, relative au modèle non paramétrique (abordé dans le cadre de variables explicatives dépendantes du temps par [MU90]) par l'expression précédente. Afin de mener le test à terme, il est nécessaire de connaître la loi de $\hat{T}^{(n)}$. Donnons dans un premier temps quelques notations qui permettent de décomposer les résidus de martingales.

Notation 2. Pour $t \in [0, 1]$ et $1 \leq i \leq n$, on note :

$$\begin{aligned}\tilde{M}^{(n)}(t; z_i) &= M_i(t) - \sum_{j=1}^n \int_0^t p_i(s; z_j) M_j(ds), \\ p_i(s, z_j) &= \frac{Y_i(s) \mathbb{1}_{\{z_i \in \chi^{(n)}(z_j)\}}}{Y^{(n)}(s; z_i)}, \\ R_i^{(n)}(t) &= \int_0^t Y_i(s) \left(\alpha(s; z_i) - \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} \right) ds, \\ \alpha^{(n)}(s; z_i) &= \sum_{j=1}^n \mathbb{1}_{\{z_j \in \chi^{(n)}(z_i)\}} Y_j(s) \alpha(s; z_j).\end{aligned}$$

On notera aussi, pour $1 \leq i, j \leq n$,

$$\begin{aligned}\delta_{ij} &= \mathbb{1}_{\{i=j\}}, \\ \bar{\delta}_{ij} &= 1 - \delta_{ij}, \\ \bar{w}_{ij} &= w_{ij} \bar{\delta}_{ij}.\end{aligned}$$

Propriété 8.2.1. Avec les notations 2, pour $t \in [0, 1]$ et $1 \leq i \leq n$, on a :

$$\hat{M}^{(n)}(t; z_i) = \tilde{M}^{(n)}(t; z_i) + R_i^{(n)}(t).$$

Remarque 8.2.2. Notons que $z_i \in \chi^{(n)}(z_j)$ est équivalent à $z_j \in \chi^{(n)}(z_i)$, puisque la partition est fixée. En particulier, on a pour $s \in [0, 1]$:

$$\frac{Y_i(s) \mathbb{1}_{\{z_i \in \chi^{(n)}(z_j)\}}}{Y^{(n)}(s; z_i)} = \frac{Y_i(s) \mathbb{1}_{\{z_j \in \chi^{(n)}(z_i)\}}}{Y^{(n)}(s; z_j)}.$$

Démonstration de la propriété 8.2.1 : Pour $t \in [0, 1]$ et $1 \leq i \leq n$, on écrit :

$$\begin{aligned}\hat{M}^{(n)}(t; z_i) &= N_i(t) - \int_0^t Y_i(s) \frac{N^{(n)}(ds; z_i)}{Y^{(n)}(s; z_i)} \\ &= M_i(t) + \int_0^t Y_i(s) \alpha(s; z_i) ds \\ &\quad - \int_0^t \frac{Y_i(s)}{Y^{(n)}(s; z_i)} \sum_{j=1}^n \mathbb{1}_{\{z_j \in \chi^{(n)}(z_i)\}} (dM_j(s) + \alpha(s; z_j) Y_j(s) ds) \\ &= M_i(t) - \sum_{j=1}^n \int_0^t \frac{Y_i(s) \mathbb{1}_{\{z_i \in \chi^{(n)}(z_j)\}}}{Y^{(n)}(s; z_i)} M_j(ds) \\ &\quad + \int_0^t Y_i(s) \left(\alpha(s; z_i) - \frac{\sum_{j=1}^n \mathbb{1}_{\{z_j \in \chi^{(n)}(z_i)\}} \alpha(s; z_j) Y_j(s)}{Y^{(n)}(s; z_i)} \right) ds,\end{aligned}$$

ce qui n'est autre que l'expression annoncée. \square

Propriété 8.2.2. Pour $t \in [0, 1]$, on a :

$$\sum_{i=1}^n \hat{M}^{(n)}(t; z_i) = 0.$$

Démonstration : Formons la somme des résidus et utilisons la remarque 8.2.2 pour le second terme :

$$\begin{aligned} \sum_{i=1}^n \hat{M}^{(n)}(t; z_i) &= \sum_{i=1}^n \left(N_i(t) - \int_0^t Y_i(s) \frac{N^{(n)}(ds; z_i)}{Y^{(n)}(s; z_i)} \right) \\ &= \sum_{i=1}^n N_i(t) - \sum_{i=1}^n \int_0^t \frac{Y_i(s) \sum_{j=1}^n \mathbb{1}_{\{z_j \in \chi^{(n)}(z_i)\}} N_j(ds)}{Y^{(n)}(s; z_i)} \\ &= \sum_{i=1}^n N_i(t) - \sum_{j=1}^n \int_0^t \frac{\sum_{i=1}^n \mathbb{1}_{\{z_i \in \chi^{(n)}(z_j)\}} Y_i(s)}{Y^{(n)}(s; z_j)} dN_j(s) \\ &= 0. \end{aligned}$$

□

On obtient ainsi une nouvelle formulation pour $\hat{T}^{(n)}$.

Corollaire 8.2.1. $\hat{T}^{(n)}$ s'exprime alors, pour $t \in [0, 1]$,

$$\hat{T}^{(n)}(t) = \sum_{i=1}^n \sum_{j=1}^n \int_0^t w_{ij} \hat{M}^{(n)}(s-; z_j) \hat{M}^{(n)}(ds; z_i).$$

Sous réserve de vérifier la normalité asymptotique de $a_n \hat{T}^{(n)}(t)$, où $(a_n)_{n \geq 1}$ est une suite normalisante, on peut utiliser :

$$\hat{T}^{(n)}(1) / \sqrt{\langle \hat{T}^{(n)} \rangle (1)},$$

pour statistique de test (voir par exemple [CJG97] dans le cadre du modèle de Cox).

Chapitre 9

Comportement asymptotique de la statistique de test

Sommaire

9.1	Définitions, notations et asymptotique de \hat{T}_n	145
9.2	Démonstration du théorème 9.1.1	146

Après avoir établi quelques notations et propriétés élémentaires sur les estimateurs et processus précédemment définis, nous donnons les conditions qui nous permettront d'obtenir la loi asymptotique de la statistique de test définie au cours du chapitre 8. Vient ensuite une seconde partie qui vise à démontrer ce résultat.

9.1 Définitions, notations et asymptotique de \hat{T}_n

ette première partie a pour objectif d'énoncer le théorème principal obtenu pour les tests d'homogénéité ; les conditions obtenues sont similaires à celles énoncées dans [MU90] et nous permettent ainsi de bénéficier des résultats relatifs au modèle issu du même article, dont le principal résultat de convergence a été rappelé dans le théorème 8.1.1.

Nous cherchons désormais à établir la normalité asymptotique de $\hat{T}^{(n)}$. Pour ce faire, nous introduisons la suite normalisante $(a_n)_{n \geq 1}$ telle que

$$a_n^{-2} = \sum_{i=1}^n \sum_{j=1}^n w_{ij}^2,$$

qui donne la vitesse de la convergence de $\hat{T}^{(n)}$. La suite $(\omega_n)_{n \geq 1}$ sera choisie de sorte que $(a_n \hat{T}^{(n)})_{n \geq 1}$ soit asymptotiquement normale. Le théorème suivant nous donne sa variance limite.

Théorème 9.1.1. *On suppose que la suite $(\omega_n)_{n \geq 1}$ satisfait $n\omega_n \rightarrow +\infty$ et $n\omega_n^2 = o(1)$, lorsque n tend vers l'infini et que les conditions suivantes sont vérifiées.*

(B0) *(condition (A2) du théorème 8.1.1)*

$$\text{Leb} \{s \in \mathcal{F}_z^{(n)}; Y^{(n)}(s; z) = 0\} = o_P(1/\sqrt{n\omega_n}),$$

lorsque $n \rightarrow +\infty$.

(B1) *(condition sur W) Il existe un entier $K \geq 2$ tel que chaque ligne de W compte moins de K termes non nuls.*

(B2a) $\sup_{z \in [0,1]} \text{Leb} \{\bar{\mathcal{F}}_z^{(n)}\} = \mathcal{O}(\omega_n)$.

(B2b) $\sup_{z \in [0,1]} \text{Card} \{1 \leq i \leq n; z_i \in \chi^{(n)}(z)\} = \mathcal{O}(n\omega_n)$.

(B3) *Il existe une fonction g mesurable, positive et bornée, définie sur le support de α telle que*

$$\sup_{z \in [0,1]} \int_{\mathcal{F}_z^{(n)}} \left(\frac{n\omega_n}{Y^{(n)}(s; z)} - g(s; z) \right)^2 ds \xrightarrow{P} 0, \text{ lorsque } n \rightarrow +\infty.$$

(B4)

$$\sup_{s \in [0,1]} \left| a_n^2 \sum_{i=1}^n \sum_{j=1}^n \bar{w}_{ij}^2 \Lambda_j(s) Y_i(s) \alpha(s; z_i) - \gamma(s) \right| \xrightarrow{P} 0,$$

où γ intégrable positive. On note $\Gamma(t) = \int_0^t \gamma(s) ds$.

Alors $a_n \hat{T}^{(n)} \xrightarrow{\mathcal{D}} \mathcal{U}$ dans $\mathcal{D}([0,1])$, où \mathcal{U} est une martingale gaussienne centrée de fonction de variance

$$\text{Var}(\mathcal{U}(t)) = \Gamma(t).$$

Si l'on fixe un seuil de confiance α , et que l'on désigne par τ une estimation de $\hat{T}^{(n)}(1)/\sqrt{\langle \hat{T}^{(n)} \rangle(1)}$, on accepte l'hypothèse H_0 si $|\tau| > u_\alpha$, où u_α est choisi tel que $\Phi(u_\alpha) = 1 - \alpha/2$ (avec Φ , la fonction de répartition d'une loi normale centrée réduite) et on la rejette sinon.

Remarque 9.1.1. *Les hypothèses du théorème 9.1.1 impliquent les hypothèses (A1) à (A3) du théorème 8.1.1, ainsi que les conditions requises sur ω_n . Une discussion sur ces hypothèses est présentée dans [MU90], qui montre que dans le cas de variables indépendantes et distribuées, les hypothèses du théorème 8.1.1 sont vérifiées. Notons ici que la condition (B3) est plus forte que celle du théorème 8.1.1.*

Démonstration : En effet, on sait que $n\omega_n$ tend vers l'infini et que $n\omega_n^2$ tend vers 0, ce qui nous donne les conditions sur la vitesse de convergence de la suite $(\omega_n)_{n \geq 1}$. De plus, la condition (B2a) implique l'hypothèse (A3). La condition (B3) implique (A2) et la condition (A1) est reprise sous le nom de condition (B0). \square

Remarque 9.1.2. *La statistique de test est dégénérée lorsque W est la matrice identité.*

La remarque suivante répertorie quelques résultats qui seront utiles lors des démonstrations qui suivent.

Remarque 9.1.3. *Par (B1), on a $n \leq a_n^{-2} \leq Kn$.*

9.2 Démonstration du théorème 9.1.1

Dans cette partie, nous démontrons le théorème 9.1.1. Pour cela, on décompose la statistique de test comme suit.

$$\begin{aligned} a_n \hat{T}^{(n)}(t) &= a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n w_{ij} \hat{M}_j^{(n)}(s-) \hat{M}_i^{(n)}(ds) \\ &= a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n w_{ij} \tilde{M}_j^{(n)}(s-) \tilde{M}_i^{(n)}(ds) \end{aligned} \quad (\text{A1})$$

$$+ a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n w_{ij} R_j^{(n)}(s-) \tilde{M}_i^{(n)}(ds) \quad (\text{A2})$$

$$+ a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n w_{ij} R_j^{(n)}(s-) R_i^{(n)}(ds) \quad (\text{A3})$$

$$+ a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n w_{ij} \tilde{M}_j^{(n)}(s-) R_i^{(n)}(ds). \quad (\text{A4})$$

Nous traitons les termes (A1) à (A4) séparément par le biais des lemmes qui suivent. Par la suite, nous noterons, $A_i(t)$ l'expression donnée en (Ai), pour $1 \leq i \leq 4$. Les résultats asymptotiques sont obtenus pour n tendant vers l'infini.

Considérons les termes (A2) et (A4) et appliquons la proposition D.0.1. On obtient, pour i et j fixés entre 1 et n , et $t \in [0, 1]$,

$$\begin{aligned} R_j^{(n)}(t)\tilde{M}_i^{(n)}(t) - R_j^{(n)}(0)\tilde{M}_i^{(n)}(0) &= \int_0^t R_j^{(n)}(s-)\tilde{M}_i^{(n)}(ds) + \int_0^t \tilde{M}_i^{(n)}(s-)R_j^{(n)}(ds) \\ &+ \sum_{0 < x \leq t} \Delta R_j^{(n)}(x)\Delta\tilde{M}_i^{(n)}(x). \end{aligned}$$

Par définitions de $\tilde{M}_i^{(n)}$, continuité de $R_i^{(n)}$ et symétrie de W , on a alors :

$$\begin{aligned} &a_n \sum_{i=1}^n \sum_{j=1}^n w_{ij} R_j^{(n)}(t)\tilde{M}_i^{(n)}(t) \\ &= a_n \sum_{i=1}^n \sum_{j=1}^n \int_0^t w_{ij} \left(R_j^{(n)}(s-)\tilde{M}_i^{(n)}(ds) + \tilde{M}_j^{(n)}(s-)R_i^{(n)}(ds) \right) \quad (A5), \end{aligned}$$

ce qui permet d'avoir une expression (notée (A5)) de la somme des termes (A2) et (A4).

Lemme 9.2.1. *Sous les hypothèses (B1) à (B4), on a*

1. $\max_{1 \leq j \leq n} \sup_{s \in [0,1]} |R_j^{(n)}(s)| = \mathcal{O}_P(\omega_n)$,
- 2.

$$\sup_{z \in [0,1]} \left| \int_0^1 \frac{Y_i(s)\alpha^{(n)}(s; z_i)}{(Y^{(n)}(s; z_i))^2} ds \right| = \mathcal{O}_P\left(\frac{1}{n\omega_n}\right).$$

Lemme 9.2.2. *On a :*

$$\sup_{t \in [0,1]} \max(|A5(t)|, |A3(t)|) \xrightarrow{P} 0.$$

Lemme 9.2.3. *On a :*

$$\sup_{t \in [0,1]} \left| A1(t) - a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right) M_i(ds) \right| \xrightarrow{P} 0.$$

Démonstration du théorème 9.1.1 :

En appliquant le lemme 9.2.2, il apparaît que, uniformément en $t \in [0, 1]$, on a :

$$a_n \hat{T}^{(n)}(t) = o_P(1) + A1(t),$$

et par le lemme 9.2.3, uniformément en $t \in [0, 1]$, on obtient :

$$a_n \hat{T}^{(n)}(t) = o_P(1) + a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right) M_i(ds).$$

Considérons $\tilde{T}^{(n)}$, le processus défini pour $t \in [0, 1]$, par :

$$\tilde{T}^{(n)}(t) = a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right) M_i(ds),$$

et $\tilde{T}_\varepsilon^{(n)}$, son processus de sauts, défini par :

$$\tilde{T}_\varepsilon^{(n)}(t) = a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right) \mathbb{1}_{\{a_n |\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i)| > \varepsilon\}} M_i(ds).$$

On vient de voir que $\tilde{T}^{(n)}$ est la partie asymptotique non négligeable de $\hat{T}^{(n)}$. On calcule donc sa loi asymptotique, qui nous permettra ainsi de déduire celle de la statistique de test. On utilise ainsi le théorème de Rebolledo, et l'on procède en deux étapes, qui correspondent aux deux hypothèses du théorème. Dans un premier temps, nous montrons que le processus de saut de $\tilde{T}^{(n)}$ converge en probabilité vers 0, puis nous déterminons la limite du processus des variations prévisibles associé à $\tilde{T}^{(n)}$.

Remarquons dans un premier temps que $\tilde{T}_\varepsilon^{(n)}$ converge en probabilité vers 0, lorsque n tend vers l'infini. En effet, pour $t \in [0, 1]$, le processus des variations prévisibles associé à $\tilde{T}_\varepsilon^{(n)}$ est donné par :

$$\left\langle \tilde{T}_\varepsilon^{(n)} \right\rangle (t) = a_n^2 \sum_{i=1}^n \int_0^t X_{n,i}^2(s) \mathbb{1}_{\{a_n |X_{n,i}(s)| > \varepsilon\}} Y_i(s) \alpha(s; z_i) ds,$$

où $X_{n,i}(s) = \sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i)$.

Or, pour $s \in [0, 1]$, on a :

$$\begin{aligned} a_n \left| \sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right| &\leq a_n \left| \sum_{j=1}^n w_{ij} M_j(s-) \right| + a_n |A(s; z_i)| \\ &\leq a_n (1 + A) \max_{1 \leq i \leq n} \sum_{j=1}^n |w_{ij}| + a_n A \\ &= \mathcal{O}(a_n) = o(1), \end{aligned}$$

où $A(\cdot, \cdot)$ est bornée par la constante A . En effet, on sait que a_n converge vers 0, par la remarque 9.1.3.

Cela nous permet de déduire que

$$\max_{1 \leq i \leq n} \sup_{t \in [0, 1]} a_n \left| \sum_{j=1}^n w_{ij} M_j(t) + A(t; z_i) \right|$$

converge presque sûrement vers 0, lorsque n tend vers l'infini, et donc que $\left\langle \tilde{T}_\varepsilon^{(n)} \right\rangle$ converge vers 0 en probabilité.

Il reste à montrer l'existence d'une fonction déterministe Γ , définie sur $[0, 1]$, telle que

$$\langle \tilde{T}^{(n)} \rangle (t) \xrightarrow[n \rightarrow +\infty]{P} \Gamma(t),$$

pour appliquer le théorème de Rebolledo. Dans un premier temps, exprimons le processus des variations prévisibles associé à $\tilde{T}^{(n)}$ en $t \in [0, 1]$:

$$\begin{aligned} \langle \tilde{T}^{(n)} \rangle (t) &= a_n^2 \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right)^2 Y_i(s) \alpha(s; z_i) ds \\ &= a_n^2 \sum_{i=1}^n \int_0^t \sum_{j=1}^n \sum_{k=1}^n w_{ij} w_{ik} M_j(s-) M_k(s-) Y_i(s) \alpha(s; z_i) ds \\ &\quad + 2a_n^2 \sum_{i=1}^n \int_0^t A(s; z_i) \sum_{j=1}^n w_{ij} M_j(s-) Y_i(s) \alpha(s; z_i) ds \\ &\quad + a_n^2 \sum_{i=1}^n \int_0^t A(s; z_i)^2 Y_i(s) \alpha(s; z_i) ds. \end{aligned}$$

En notant que, pour $1 \leq i \leq n$, on a presque sûrement

$$N_i(t-) Y_i(t) = 0, \quad (9.2)$$

$$Y_i(t) \Lambda_i(t) = -Y_i(t) M_i(t-) = Y_i(t) A_i(t), \quad (9.3)$$

on aboutit à l'expression suivante :

$$\langle \tilde{T}^{(n)} \rangle (t) = a_n^2 \sum_{i=1}^n \int_0^t \sum_{j=1}^n \sum_{k=1}^n \bar{\delta}_{jk} w_{ij} w_{ik} M_j(s-) M_k(s-) Y_i(s) \alpha(s; z_i) ds \quad (9.4)$$

$$+ a_n^2 \sum_{i=1}^n \int_0^t \sum_{j=1}^n \bar{w}_{ij}^2 (M_j^2(s-) - \Lambda_j(s)) Y_i(s) \alpha(s; z_i) ds \quad (9.5)$$

$$+ 2a_n^2 \sum_{i=1}^n \int_0^t A_i(s) \sum_{j=1}^n \bar{w}_{ij} M_j(s-) Y_i(s) \alpha(s; z_i) ds \quad (9.6)$$

$$+ a_n^2 \sum_{i=1}^n \int_0^t \sum_{j=1}^n \bar{w}_{ij}^2 \Lambda_j(s) Y_i(s) \alpha(s; z_i) ds. \quad (9.7)$$

Par l'hypothèse (B4), on sait que (9.7) converge en probabilité vers Γ . On termine la preuve en montrant que les termes (9.4), (9.5) et (9.6) convergent vers 0 en probabilité, uniformément en t . Pour cela, on utilise la proposition D.0.2 démontrée dans [BC01].

Le terme (9.5) obtenu est semblable au terme (3b) de [BC01]. On remarque dans un premier temps que l'on a :

$$\sup_{t \in [0,1]} \left| a_n^2 \sum_{i=1}^n \sum_{j=1}^n w_{ij}^2 \bar{\delta}_{ij} (M_j^2(t-) - \Lambda_j(t)) Y_i(t) \alpha(t; z_i) \right| = \mathcal{O}(a_n^2 n) = \mathcal{O}(1).$$

Ainsi la première hypothèse de la proposition D.0.2 est satisfaite. De plus, pour $t \in [0, 1]$, il existe une constante positive C , telle que :

$$\begin{aligned} & \mathbb{E} \left(a_n^2 \sum_{i=1}^n \sum_{j=1}^n \overline{w_{ij}^2} (M_j^2(t-) - \Lambda_j(t)) Y_i(t) \alpha(t; z_i) \right)^2 \\ & \leq C a_n^4 \left(\sum_{i=1}^n \sum_{j=1}^n w_{ij}^2 + \sum_{i=1}^n \sum_{j=1}^n \sum_{k=1}^n |w_{ij} w_{ik}| \right) \\ & = \mathcal{O}(n a_n^4), \end{aligned}$$

par l'hypothèse (B1). On sait que la suite $(a_n)_{n \geq 1}$ converge vers 0 par la remarque 9.1.3, et que $n a_n^2$ est borné, ce qui nous permet de déduire que le terme précédent converge vers 0 et nous donne le fait que le terme (9.5) converge en probabilité vers 0, uniformément en t .

Les termes (9.4), (9.5) et (9.6) sont semblables, puisque $\tilde{M}_j^{(n)} = M_j^2 - \Lambda_j$ et M_j sont des martingales, et α et $A\alpha$ sont bornées uniformément sur $[0, 1]^2$. On applique alors la même méthode que celle utilisée pour le terme (3b) dans [BC01]. Cela nous permet ainsi de conclure. \square

Démonstration du lemme 9.2.1 :

1. Par définition, pour $t \in [0, 1]$, on a :

$$R_i^{(n)}(t) = \int_0^t Y_i(s) \left(\alpha(s; z_i) - \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} \right) ds.$$

On sait que α et $\alpha^{(n)}$ s'annulent en dehors de $\mathcal{F}_{z_i}^{(n)} \cup \overline{\mathcal{F}_{z_i}^{(n)}}$, donc on peut écrire

$$\begin{aligned} |R_i^{(n)}(t)| & \leq \int_{\mathcal{F}_{z_i}^{(n)}} Y_i(s) \left| \alpha(s; z_i) - \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} \right| ds \\ & \quad + \int_{\overline{\mathcal{F}_{z_i}^{(n)}}} Y_i(s) \left| \alpha(s; z_i) - \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} \right| ds. \end{aligned}$$

D'après [MU90], on a sur $\mathcal{F}_{z_i}^{(n)}$ (voir remarque 8.1.2), uniformément en i ,

$$\alpha^{(n)}(s; z_i) = (\alpha(s; z_i) + \mathcal{O}_P(\omega_n)) Y^{(n)}(s; z_i).$$

Par ailleurs, la seconde intégrale est un $\mathcal{O}(\overline{\mathcal{F}_{z_i}^{(n)}}) = \mathcal{O}(\omega_n)$, grâce à l'hypothèse (B2a) concernant $\overline{\mathcal{F}_{z_i}^{(n)}}$, et par le fait que α est lipschitzienne donc bornée sur un compact. On en déduit la relation suivante

$$\max_{1 \leq i \leq n} \sup_{t \in [0, \tau]} |R_i^{(n)}(t)| = \mathcal{O}_P(\omega_n),$$

qui est le résultat souhaité.

2. Soit D_i définie par :

$$D_i = \int_0^1 \frac{Y_i(s) \alpha^{(n)}(s; z_i)}{(Y^{(n)}(s; z_i))^2} ds$$

On a alors :

$$D_i = \int_0^1 \frac{Y_i(s)}{Y^{(n)}(s; z_i)} \left(\frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} - \alpha(s; z_i) \right) ds$$

$$+ \int_0^1 \frac{Y_i(s)}{Y^{(n)}(s; z_i)} \alpha(s; z_i) ds,$$

$$|D_i| \leq \int_{\mathcal{F}_{z_i}^{(n)}} \frac{Y_i(s)}{Y^{(n)}(s; z_i)} \left| \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} - \alpha(s; z_i) \right| ds \quad (\text{H1})$$

$$+ \int_{\overline{\mathcal{F}}_{z_i}^{(n)}} \frac{Y_i(s)}{Y^{(n)}(s; z_i)} \left| \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} - \alpha(s; z_i) \right| ds \quad (\text{H2})$$

$$+ \int_{\mathcal{F}_{z_i}^{(n)}} \frac{Y_i(s)}{Y^{(n)}(s; z_i)} \alpha(s; z_i) ds \quad (\text{H3})$$

$$+ \int_{\overline{\mathcal{F}}_{z_i}^{(n)}} \frac{Y_i(s)}{Y^{(n)}(s; z_i)} \alpha(s; z_i) ds. \quad (\text{H4})$$

Traisons les quatre termes séparément et montrons que chacun d'entre eux est au moins un $\mathcal{O}_P((n\omega_n)^{-1})$. Les résultats qui suivent sont uniformes en i , on a, sur $\mathcal{F}_{z_i}^{(n)}$:

$$\left| \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} - \alpha(s; z_i) \right| = \mathcal{O}_P(\omega_n),$$

par la remarque 8.1.2. Cela nous donne ainsi que (H1) est un $\mathcal{O}(\omega_n)$. Le terme (H2) est un $\mathcal{O}\left(\text{Leb}\left(\overline{\mathcal{F}}_{z_i}^{(n)}\right)\right) = \mathcal{O}(\omega_n)$. Ensuite, on majore le terme (H3) par

$$\begin{aligned} & C \left(\frac{1}{n\omega_n} \int_{\mathcal{F}_{z_i}^{(n)}} \left| \frac{n\omega_n}{Y^{(n)}(s; z_i)} - g(s; z_i) \right| ds + \int_{\mathcal{F}_{z_i}^{(n)}} \frac{g(s; z_i)}{n\omega_n} ds \right) \\ &= o_p\left(\frac{1}{n\omega_n}\right) + \mathcal{O}\left(\frac{1}{n\omega_n}\right), \end{aligned}$$

où C est une constante positive.

Enfin, le terme (H4) est un $\mathcal{O}\left(\text{Leb}\left(\overline{\mathcal{F}}_{z_i}^{(n)}\right)\right) = \mathcal{O}(\omega_n)$, et en regroupant tous les termes, on obtient finalement $D_i = \mathcal{O}_P((n\omega_n)^{-1})$, car $\omega_n = o(n^{-1}\omega_n^{-1})$ (en effet, $\omega_n^2 n$ tend vers 0, par hypothèse). Le résultat est uniforme en z , puisqu'aucune majoration ne dépend de z . \square

Démonstration du lemme 9.2.2. On écrit une majoration de (A5) :

$$\begin{aligned}
a_n \left| \sum_{i=1}^n \sum_{j=1}^n w_{ij} \tilde{M}_j^{(n)}(t) R_i^{(n)}(t) \right| &\leq a_n \max_{1 \leq i \leq n} \sup_{t \in [0,1]} |R_i^{(n)}(t)| \sum_{i=1}^n \sum_{j=1}^n |w_{ij}| |\tilde{M}_j^{(n)}(t)| \\
&\leq a_n \max_{1 \leq i \leq n} \sup_{t \in [0,1]} |R_i^{(n)}(t)| \\
&\times \sum_{k=1}^n \int_0^t \sum_{j=1}^n \left(\sum_{i=1}^n |w_{ij}| \right) \frac{Y_j(u) \mathbb{1}_{\{z_j \in \chi^{(n)}(z_k)\}}}{Y^{(n)}(u; z_k)} |M_k(du)| \\
&+ a_n \max_{1 \leq i \leq n} \sup_{t \in [0,1]} |R_i^{(n)}(t)| \sum_{i=1}^n \sum_{j=1}^n |w_{ij}| |M_j(t)| \\
&\leq \mathcal{O}_P(na_n\omega_n) + \mathcal{O}_P \left(a_n \max_{1 \leq i \leq n} \sup_{t \in [0,1]} |R_i^{(n)}(t)| \right) \\
&\times \sum_{k=1}^n \int_0^t \frac{\sum_{j=1}^n Y_j(u) \mathbb{1}_{\{z_j \in \chi^{(n)}(z_k)\}}}{Y^{(n)}(u; z_k)} |M_k(du)| \\
&= \mathcal{O}_P(na_n\omega_n) = o_P(1),
\end{aligned}$$

qui nous permet de déduire que (A5) converge en probabilité vers 0, uniformément sur $[0, \tau]$.

Il nous reste désormais à traiter le terme (A3). On utilise pour cela le lemme 9.2.1, la remarque 9.1.3 ainsi que le calcul effectué précédemment pour majorer (A5) et on obtient :

$$\begin{aligned}
|A3(t)| &\leq a_n \sum_{i=1}^n \left| \int_0^t \sum_{j=1}^n w_{ij} R_j^{(n)}(s-) R_i^{(n)}(ds) \right| \\
&\leq a_n \max_{1 \leq i \leq n} \sum_{j=1}^n |w_{ij}| \max_{1 \leq j \leq n} \sup_{s \in [0,1]} |R_j^{(n)}(s)| \sum_{i=1}^n \int_0^t |R_i^{(n)}(ds)|.
\end{aligned}$$

D'après la formulation de $R_i^{(n)}$, on a

$$\frac{dR_i^{(n)}}{ds}(s) = Y_i(s) \left(\alpha(s; z_i) - \frac{\alpha^{(n)}(s; z_i)}{Y^{(n)}(s; z_i)} \right).$$

Il vient alors

$$\sup_{t \in [0,1]} |A3(t)| = \mathcal{O}_P(a_n n \omega_n) = o_P(1).$$

Cela achève la démonstration de ce lemme. \square

Preuve du lemme 9.2.3. On décompose $A1(t)$ comme suit

$$\begin{aligned} A1(t) &= a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n w_{ij} \tilde{M}_j^{(n)}(s-) \tilde{M}_i^{(n)}(ds) \\ &= a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) - \sum_{j=1}^n \sum_{l=1}^n w_{lj} p_l(s; z_i) M_j(s-) \right) M_i(ds) \quad (I1) \end{aligned}$$

$$+ a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n \sum_{l=1}^n w_{ij} \int_0^{s-} p_j(u; z_l) M_l(du) \sum_{k=1}^n p_i(s; z_k) M_k(ds) \quad (I2)$$

$$- a_n \sum_{i=1}^n \int_0^t \sum_{j=1}^n w_{ij} \sum_{l=1}^n \int_0^{s-} p_j(u; z_l) M_l(du) M_i(ds) \quad (I3)$$

Pour $i = 1, 2$ et 3 , on note $I_i(t)$ l'expression donnée en (Ii). Pour montrer ce résultat, deux étapes sont nécessaires, elles sont résumées dans les lemmes 9.2.4 et 9.2.6. Le lemme 9.2.5 n'est qu'une étape pour prouver le résultat du lemme 9.2.6.

Lemme 9.2.4. On a :

$$\sup_{t \in [0,1]} \max(|I2(t)|, |I3(t)|) \xrightarrow{P} 0.$$

Lemme 9.2.5. Le processus $A0$ défini pour $t \in [0, 1]$ par :

$$A0(t) = a_n \sum_{i=1}^n \int_0^t \sum_{l=1}^n \sum_{j=1}^n \bar{w}_{lj} M_j(s-) p_l(s; z_i) M_i(ds)$$

converge en probabilité vers 0, uniformément sur $[0, 1]$.

Lemme 9.2.6. On a :

$$\sup_{t \in [0,1]} \left| I1(t) - a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right) M_i(ds) \right| \xrightarrow{P} 0.$$

Par le lemme 9.2.4, on sait qu'il suffit de ne considérer que le terme (I1) pour le calcul de la limite uniforme sur $[0, 1]$ de $A1$. Par ailleurs, par le lemme 9.2.6, on obtient :

$$\sup_{t \in [0,1]} \left| A1(t) - a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right) M_i(ds) \right| \xrightarrow{P} 0,$$

ce qui n'est autre que le résultat précédemment annoncé. \square

Démonstration du lemme 9.2.4. On a par l'inégalité de Cauchy-Schwarz

$$\begin{aligned}
\langle I3 \rangle (1) &\leq C \int_0^1 a_n^2 \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} \int_0^{s-} \frac{Y_j(u)M^{(n)}(du; z_j)}{Y^{(n)}(u, z_j)} \right)^2 ds \\
&\leq C' \int_0^1 a_n^2 \sum_{i=1}^n \sum_{j=1}^n |w_{ij}| \left(\int_0^{s-} \frac{Y_j(u)M^{(n)}(du; z_j)}{Y^{(n)}(u, z_j)} \right)^2 ds \\
&\leq C' a_n^2 \sum_{i=1}^n \sum_{j=1}^n |w_{ij}| \sup_{0 \leq s \leq 1} \left(\int_0^s \frac{Y_j(u)M^{(n)}(du; z_j)}{Y^{(n)}(u, z_j)} \right)^2,
\end{aligned}$$

où C et C' sont des constantes positives. Pour $1 \leq j \leq n$, les $\left(\int_0^s \frac{Y_j(u)M^{(n)}(du; z_j)}{Y^{(n)}(u, z_j)} \right)^2$ sont des sous-martingales, on peut donc appliquer l'inégalité de Lengart. La sous-martingale

$$a_n^2 \sum_{i=1}^n \sum_{j=1}^n |w_{ij}| \left(\int_0^s \frac{Y_j(u)M^{(n)}(du; z_j)}{Y^{(n)}(u, z_j)} \right)^2$$

est compensée par :

$$\begin{aligned}
K_n(s) &= a_n^2 \sum_{i=1}^n \sum_{j=1}^n |w_{ij}| \int_0^s \frac{Y_j(u)\alpha^{(n)}(u; z_j)du}{(Y^{(n)}(u, z_j))^2} \\
&\leq K a_n^2 n \sup_{0 \leq z \leq 1} \int_0^1 \frac{\mathbb{1}_{\{Y^{(n)}(u, z) > 0\}} \alpha^{(n)}(u; z)du}{(Y^{(n)}(u, z))^2} = O_P(1/(n\omega_n)) = o_P(1),
\end{aligned}$$

par le fait que $a_n^2 = O(1/n)$ et le lemme 9.2.1.

Par ailleurs, on peut voir que :

$$\langle I2 \rangle (1) = \int_0^1 a_n^2 \sum_{l=1}^n Y_l(s)\alpha(s, z_l) \left(\sum_{i=1}^n p_i(s; z_l) \sum_{j=1}^n w_{ij} \int_0^{s-} \frac{Y_j(u)M^{(n)}(du; z_j)}{Y^{(n)}(u; z_j)} \right)^2 ds.$$

En remplaçant $p_i(s; z_l)$ par $\sqrt{p_i(s; z_l)}\sqrt{p_i(s; z_l)}$ puis en utilisant l'inégalité de Cauchy-Schwarz, on obtient alors :

$$\langle I2 \rangle (1) \leq C \int_0^1 a_n^2 \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} \int_0^{s-} \frac{Y_j(u)M^{(n)}(du; z_j)}{Y^{(n)}(u; z_j)} \right)^2 \left(\sum_{l=1}^n Y_l(s)p_i(s; z_l) \right) ds,$$

où C est une constante positive. En notant que $\sum_{l=1}^n Y_l(s)p_i(s; z_l) = 1$, il est possible d'utiliser les résultats obtenus pour (I3) pour conclure. \square

Démonstration du lemme 9.2.5. Pour $t \in [0, 1]$, on a, pour une constante C

positive :

$$\langle A0 \rangle (t) \tag{9.10}$$

$$= a_n^2 \sum_{i=1}^n \int_0^t \left(\sum_{l=1}^n \sum_{j=1}^n \bar{w}_{lj} M_j(s-) p_l(s; z_i) \right)^2 Y_i(s) \alpha(s; z_i) ds$$

$$\leq \frac{C a_n^2}{(n \omega_n)^2} \sum_{i=1}^n \int_0^t h(s; z_i) g(s; z_i)^2 Y_i(s) ds \tag{9.11}$$

$$+ C \frac{a_n^2}{n^2 \omega_n^2} \sum_{i=1}^n \int_0^t h(s; z_i) \left(\frac{n^2 \omega_n^2}{Y^{(n)}(s; z_i)^2} - g(s; z_i)^2 \right) Y_i(s) \alpha(s; z_i) ds, \tag{9.12}$$

où

$$h(s; z_i) = \left(\sum_{l=1}^n \sum_{j=1}^n \bar{w}_{lj} Y_l(s) \mathbb{1}_{\{z_l \in \chi^{(n)}(z_i)\}} M_j(s-) \right)^2.$$

On remarque que (9.11) peut se majorer par un terme de la forme suivante (on sait que g est une fonction bornée)

$$C \frac{a_n^2}{(n \omega_n)^2} \int_0^1 X_n(s) ds,$$

où

$$X_n(s) = \sum_{i=1}^n Y_i(s) \left(\sum_{l=1}^n \sum_{j=1}^n \bar{w}_{lj} M_j(s-) Y_l(s) \mathbb{1}_{\{z_l \in \chi^{(n)}(z_i)\}} \right)^2.$$

On doit alors montrer que $Y_n = a_n^2 n^{-2} \omega_n^{-2} X_n$ est uniformément, presque sûrement bornée et que $\sup_{s \in [0,1]} \mathbb{E} |Y_n(s)| \rightarrow 0$. Pour $s \in [0, 1]$, exprimons $\mathbb{E} |X_n(s)|$ (s est désormais omis pour alléger l'écriture) :

$$\mathbb{E} |X_n| = \mathbb{E} \left(\sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{k=1}^n \sum_{m \in \chi^{(n)}(z_i)} \sum_{j=1}^n Y_i \bar{w}_{lj} \bar{w}_{mk} Y_l Y_m M_j M_k \right),$$

où la notation « $l \in \chi^{(n)}(z_i)$ » désigne l'ensemble des entiers l pour lesquels z_l est dans l'ensemble $\chi^{(n)}(z_i)$. Examinons la somme portant sur l'indice j . Par indépendance de M_j par rapport aux variables aléatoires Y_i, Y_l, Y_m et M_k , on remarque que cette somme est nulle pour $j \notin \{k, m, l, i\}$, puisque M_j est d'espérance nulle. En outre, les termes \bar{w}_{lj} et \bar{w}_{mk} nous assurent qu'on a simultanément m et k , ainsi que l et j distincts. On distingue alors trois composantes d'intersection vide susceptibles d'être non nulles pour ce terme :

$$\begin{aligned} j &= k; \\ j &= i, j \neq k; \\ j &= m, j \neq k, j \neq i. \end{aligned}$$

En utilisant (9.2) et (9.3), on obtient :

$$\begin{aligned}
\mathbb{E}|X_n(s)| &= \mathbb{E} \left(\sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{k=1}^n \sum_{m \in \chi^{(n)}(z_i)} Y_i \bar{w}_{lk} \bar{w}_{mk} Y_l Y_m M_k^2 \right) \\
&+ \mathbb{E} \left(\sum_{k=1}^n \sum_{i=1, i \neq k}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{m \in \chi^{(n)}(z_i)} Y_i \bar{w}_{li} \bar{w}_{mk} Y_l Y_m (-\Lambda_i) M_k \right) \\
&+ \mathbb{E} \left(\sum_{k=1}^n \sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{m \in \chi^{(n)}(z_i), m \neq i, k} Y_i \bar{w}_{lm} \bar{w}_{mk} Y_l Y_m (-\Lambda_m) M_k \right).
\end{aligned}$$

D'autres raisonnements similaires permettent d'obtenir :

$$\begin{aligned}
\mathbb{E}|X_n(s)| &= \mathbb{E} \left(\sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{k=1}^n \sum_{m \in \chi^{(n)}(z_i)} Y_i \bar{w}_{lk} \bar{w}_{mk} Y_l Y_m M_k^2 \right) \\
&+ \mathbb{E} \left(\sum_{i=1}^n \sum_{l \in \chi^{(n)}, l \neq i(z_i)} \sum_{m \in \chi^{(n)}(z_i)} Y_i \bar{w}_{li} \bar{w}_{ml} Y_m (-\Lambda_i) (-\Lambda_l) \right) \\
&+ \mathbb{E} \left(\sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{m \in \chi^{(n)}(z_i), m \neq i} \bar{w}_{lm} \bar{w}_{mi} Y_l Y_m (-\Lambda_m) (-\Lambda_i) \right) \\
&+ \mathbb{E} \left(\sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{m \in \chi^{(n)}(z_i), m \neq i, l} Y_i \bar{w}_{lm} \bar{w}_{ml} Y_m (-\Lambda_m) (-\Lambda_l) \right) \\
&\leq C \sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} \sum_{k=1}^n |w_{lk}| \max_{1 \leq k \leq n} \sum_{m=1}^n |w_{mk}| \\
&+ C \sum_{i=1}^n \sum_{l \in \chi^{(n)}(z_i)} |w_{li}| \max_{1 \leq l \leq n} \sum_{m=1}^n |w_{ml}| \\
&+ C \sum_{i=1}^n \sum_{m \in \chi^{(n)}(z_i)} |w_{mi}| \max_{1 \leq m \leq n} \sum_{l=1}^n |w_{ml}| \\
&+ Cn \max_{1 \leq i \leq n} |\chi^{(n)}(z_i)| \max_{1 \leq l \leq n} \sum_{m=1}^n |w_{ml}|,
\end{aligned}$$

ce qui nous donne :

$$\mathbb{E}|X_n(s)| = C'n \left(\max_{1 \leq i \leq n} |\chi^{(n)}(z_i)| + 1 \right),$$

où C et C' sont des constantes positives.

Ainsi, nous obtenons :

$$\mathbb{E}|Y_n(s)| = \mathbb{E} \left| \frac{a_n^2}{n^2\omega_n^2} X_n(s) \right| = \mathcal{O} \left(\frac{a_n^2}{n\omega_n^2} \right) = o(1).$$

On peut ainsi affirmer par la proposition D.0.2 que (9.11) converge en probabilité vers 0, uniformément sur $[0, 1]$.

Au sein du terme (9.12), on sépare les intégrales sur $[0, 1]$ à l'aide de $\mathcal{F}_{z_i}^{(n)}$ et $\overline{\mathcal{F}}_{z_i}^{(n)}$ et l'on obtient :

$$a_n^2 \sum_{i=1}^n \int_{\overline{\mathcal{F}}_{z_i}^{(n)}} h(s; z_i) \frac{1}{Y^{(n)}(s; z_i)^2} Y_i(s) \alpha(s; z_i) ds \quad (9.13)$$

$$- \frac{a_n^2}{n^2\omega_n^2} \sum_{i=1}^n \int_{\overline{\mathcal{F}}_{z_i}^{(n)}} h(s; z_i) g(s; z_i)^2 Y_i(s) \alpha(s; z_i) ds \quad (9.14)$$

$$+ \frac{a_n^2}{n^2\omega_n^2} \sum_{i=1}^n \int_{\mathcal{F}_{z_i}^{(n)}} h(s; z_i) \left(\frac{n^2\omega_n^2}{Y^{(n)}(s; z_i)^2} - g(s; z_i)^2 \right) Y_i(s) \alpha(s; z_i) ds. \quad (9.15)$$

Dans un premier temps, on remarque que

$$\sup_{z \in [0,1]} h(s; z) = \mathcal{O} \left(\sup_{z \in [0,1]} \text{Card} \{1 \leq i \leq n; z_i \in \chi^{(n)}(z)\} \right)^2 = \mathcal{O}(n^2\omega_n^2),$$

ce qui nous permet de déduire que (9.14) tend vers 0, puisqu'il est un $\mathcal{O}(\omega_n)$. Ensuite, on remarque que

$$\frac{h(s; z_i)}{Y^{(n)}(s; z_i)^2} = \mathcal{O}(1),$$

pour obtenir la majoration par $Cna_n^2\omega_n = o(1)$ du terme (9.13) (C désigne une constante positive).

Enfin, pour une autre constante positive C , on majore (9.15) par :

$$Cn^3\omega_n^2 \frac{a_n^2}{n^2\omega_n^2} \sup_{z \in [0,1]} \int_{\mathcal{F}_z^{(n)}} \left| \frac{n^2\omega_n^2}{Y^{(n)}(s; z)^2} - g(s; z)^2 \right| Y_i(s) ds,$$

qui tend vers 0, par l'hypothèse (B3). Cela achève la démonstration de ce lemme. \square

Démonstration du lemme 9.2.6. Calculons la différence suivante :

$$\begin{aligned} & I1(t) - a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n w_{ij} M_j(s-) + A(s; z_i) \right) M_i(ds) \\ &= -a_n \sum_{i=1}^n \int_0^t \left(\sum_{j=1}^n \sum_{l=1}^n w_{lj} p_l(s; z_i) M_j(s-) + A(s; z_i) \right) M_i(ds). \end{aligned}$$

En utilisant le lemme 9.2.5, il nous reste à démontrer que

$$-a_n \sum_{i=1}^n \int_0^t \sum_{l=1}^n M_l(s-) p_l(s; z_i) M_i(ds) - a_n \sum_{i=1}^n \int_0^t A(s; z_i) M_i(ds)$$

converge uniformément vers 0, en probabilité.

Or, en utilisant les résultats (9.2) et (9.3), on obtient

$$\begin{aligned} & a_n \sum_{i=1}^n \int_0^t \left(\sum_{l=1}^n p_l(s; z_i) M_l(s-) + A(s; z_i) \right) M_i(ds) \\ &= a_n \sum_{i=1}^n \int_0^t \left(- \sum_{l=1}^n \frac{A(s; z_l) Y_l(s) \mathbb{1}_{\{z_l \in \chi^{(n)}(z_i)\}}}{Y^{(n)}(s; z_i)} + A(s; z_i) \right) M_i(ds) \\ &\leq a_n \sum_{i=1}^n \int_0^t \sup_{z \in \chi^{(n)}(z_i)} |A(s; z) - A(s; z_i)| |M_i(ds)| \\ &= \mathcal{O}(a_n n \omega_n) = o(1), \end{aligned}$$

en utilisant le fait que α est lipschitzienne. □

Annexe D

Quelques rappels de résultats

Nous rappelons dans un premier temps la formule d'intégration par parties que l'on peut trouver par exemple dans [FH90].

Proposition D.0.1. *Soient $F : \mathbb{R}_+ \rightarrow \mathbb{R}$ et $G : \mathbb{R}_+ \rightarrow \mathbb{R}$, des fonctions continues à droite à variation bornée sur tout intervalle fini. Alors, on a :*

$$F(t)G(t) - F(0)G(0) = \int_0^t F(x^-)dG(x) + \int_0^t G(x)dF(x),$$

et

$$F(t)G(t) - F(0)G(0) = \int_0^t F(x^-)dG(x) + \int_0^t G(x^-)dF(x) + \sum_{0 < x \leq t} \Delta F(x)\Delta G(x).$$

Le résultat qui suit est issu de [BC01].

Proposition D.0.2. *Soit $(X_n)_{n \geq 1}$ une suite de processus définis sur le même espace probabilisé. On considère les conditions suivantes.*

- (i) *pour tout n , $\sup_{s \in [0,1]} |X_n(s)|$ est presque sûrement bornée,*
- (ii) *$\sup_{s \in [0,1]} \mathbb{E}X_n(s)^2$ tend vers 0, lorsque $n \rightarrow +\infty$,*
- (iii) *$\sup_{s \in [0,1]} \mathbb{E}|X_n(s)|$ tend vers 0, lorsque $n \rightarrow +\infty$.*

Si la condition (i) ainsi que l'une des conditions (ii) et (iii) sont vérifiées, alors, on a

$$\sup_{t \in [0,1]} \left| \int_0^t X_n(s)ds \right| \xrightarrow{P} 0.$$

Corollaire D.0.1 (Corollaire de l'inégalité de Lenglart). *Soit M , une martingale de carré intégrable. On suppose que $\langle M \rangle$ converge vers 0 en probabilité. Alors, $\sup_{t \in [0,1]} |M(t)|$ converge vers 0 en probabilité.*

Démonstration : Il s'agit d'une application directe de l'inégalité de Lenglart, que l'on peut par exemple trouver dans [AG82]. \square

Conclusions et perspectives

L'estimation séquentielle a pour but de fournir des estimateurs optimaux au sens du contrôle de certains critères, comme les régions de confiance, l'erreur quadratique moyenne, etc. (voir par exemple [Gho91, DG00]). L'avantage majeur de ces méthodes réside en général dans l'interprétation intuitive qu'offrent les critères d'arrêt et dans la simplicité de leur mise en œuvre. Mais ces méthodes comportent des limites évidentes. En effet, comme on a pu le voir au cours de la première partie de cette thèse, pour la règle d'arrêt basée sur les intervalles de confiance de longueur fixée, l'asymptotique sur la taille n de l'échantillon est transformée en une asymptotique sur la longueur d de l'intervalle de confiance. Il n'y a alors pas, outre les simulations, de critère objectif pour affirmer que la valeur fixée d assure d'être proche des conditions asymptotiques.

Au cours de la première partie de ce mémoire, nous avons étudié une classe de variables d'arrêt, notée \mathcal{A} , pour laquelle nous avons établi des propriétés de convergence, mais aussi étendu les propriétés de l'estimateur du paramètre de régression du modèle à risques proportionnels de Cox au cas de données issues d'une règle d'arrêt élément de \mathcal{A} . En effet, les propriétés asymptotiques établies dans [AG82] ne se prolongent pas de façon immédiate au cas de données issues d'une procédure séquentielle d'arrêt, puisque cette dernière fait intervenir une taille d'échantillon aléatoire modélisée par un temps d'arrêt. Notre travail a ainsi consisté dans un premier temps à définir ces critères d'arrêt et les caractériser par l'appartenance à une famille de procédures d'arrêt satisfaisant des conditions de régularité. Par la suite, nous avons établi un théorème général pour les estimateurs du paramètre euclidien que nous avons ensuite particularisé au cas du modèle à risques proportionnels de Cox [Cox72], en dressant une liste d'hypothèses sur ce modèle. La vérification de ces hypothèses sous les conditions habituellement rencontrées pour le modèle de Cox occupe une grande partie de ce travail, elle fait majoritairement intervenir les processus de comptage (voir par exemple [FH90]), ainsi que les inégalités maximales pour les processus empiriques rappelées en annexe, mais aussi des techniques utilisant les propriétés des martingales. Pour plus de simplicité, ces résultats sont de prime abord obtenus dans le cas d'un paramètre de régression unidimensionnel, la généralisation au cadre multidimensionnel est abordée par la suite. Afin de pouvoir juger de la variabilité de la variable d'arrêt définissant la taille de l'échantillon, nous avons établi ensuite un résultat pour sa normalité asymptotique, qui fait l'objet du chapitre 5. Les démonstrations de ce résultat font très largement intervenir les techniques précédemment utilisées et nécessitent quelques hypothèses supplémentaires.

Au cours des chapitres 3 à 5, nous avons ainsi pu observer l'importance de la convergence complète. Elle permet le transfert de la continuité uniforme en probabilité (condition d'Anscombe) entre deux suites de variables aléatoires équivalentes au sens de la convergence complète, alors que la convergence presque sûre ne suffit pas. La convergence complète apparaît donc comme l'outil adapté au type de résultat caractérisé par la convergence de variables aléatoires pour des tailles d'échantillon caractérisées par un temps d'arrêt.

Cependant, il est pertinent de souligner que les résultats énoncés au chapitre 4, qui concernent l'extension des propriétés des estimateurs relatifs au modèle de Cox au cas de tailles d'échantillon définies par un critère séquentiel d'arrêt, ne se restreignent pas au cas de la procédure d'arrêt basée sur l'amplitude de l'intervalle de confiance. Ainsi, si l'on désire mener une étude séquentielle basée sur une autre règle d'arrêt, il suffit de montrer son appartenance à \mathcal{A} , c'est-à-dire l'existence et la régularité, ce qui élargit le champ d'applications de nos résultats.

Par ailleurs, les inégalités maximales adaptées de [Pol90] (voir Annexe A) se révèlent être un moyen performant pour aboutir à la convergence complète de processus empiriques.

Enfin, des simulations numériques nous ont permis d'illustrer l'ensemble de cette première partie. Elle permettent en outre d'observer les limites des méthodes séquentielles basées sur une règle d'arrêt définie par une amplitude d'intervalles de confiance. Nous avons ainsi pu constater un temps de calcul assez grand pour des grandes tailles d'échantillon ; mais aussi un algorithme d'arrêt qui converge lentement pour de faibles valeurs de d (qui est l'amplitude maximale de l'intervalle de confiance, pour un niveau de confiance) arbitrairement petites. Un point important concernant ces résultats est le fait qu'une amplitude égale à 1 suffit généralement à obtenir des estimations convenables pour le paramètre de régression.

Ainsi, nous avons défini le seuil d'arrêt comme fonction de l'estimateur du paramètre de régression, ce qui revient à considérer l'amplitude de l'intervalle de confiance comme une erreur relative sur l'estimation de β_0 . Nous avons pu alors constater une vitesse de convergence plus élevée. Cela ajoute de l'intérêt à l'estimation séquentielle qui peut s'avérer utile en amont de méthodes plus difficiles à mettre en œuvre, telles que la sélection de modèles [Let00].

Nos résultats se placent dans le cas d'un intervalle d'étude du type $[0, \tau]$, il serait intéressant de considérer une généralisation dans le cas où \mathbb{R}_+ définit l'intervalle d'étude. [AG82] ont déjà abordé ce point pour les propriétés de convergence et de normalité asymptotique de l'estimateur du paramètre de régression.

Par ailleurs, nous avons adapté un théorème issu de [DG00] au cas d'estimateurs définis par la maximisation d'une fonction d'estimation dans le cas semi-paramétrique. Il pourrait être intéressant de considérer d'autres modèles caractérisés par de tels M -estimateurs pour lesquels une procédure séquentielle d'arrêt est adaptée (*i.e.* qui fassent intervenir un type de données semblable à celui étudié lors du paragraphe 1.4 concernant les procédures séquentielles) et ainsi d'étudier les hypothèses sous lesquelles les conditions (C1) à (C4) sont vérifiées. Citons par exemple les

modèles à risques additifs (que l'on peut trouver dans [LY94]), additifs multiplicatifs [LY95a] ou encore les modèles de vie accélérée [LY95b].

En outre, nous avons vu que les résultats présentés lors des chapitres 3 et 4 font l'objet d'une extension au cas de paramètre de régression multidimensionnel, des simulations numériques pourraient par la suite corroborer ces résultats et nous pourrions également étudier l'influence de la dimension du paramètre pour en déduire des heuristiques sur le comportement des estimations séquentielles lorsque la dimension de l'espace du paramètre de régression croît.

Les résultats du chapitre 5 concernent l'étude de la variabilité de la variable d'arrêt définissant la taille de l'échantillon et permettent d'obtenir, sous de nouvelles conditions, sa normalité asymptotique. L'étude menée se place dans le cas où le paramètre de régression est réel, une étude dans le cas multidimensionnel pourrait être menée. Elle ne nécessite *a priori* pas de difficulté majeure, puisque les résultats du chapitre 5 sont essentiellement basés sur le comportement de l'estimateur de la variance asymptotique. Il serait en outre intéressant de mener des études numériques pour étudier la variabilité de N_d de façon qualitative.

Concernant cette première partie, une étude du comportement asymptotique pour des tailles d'échantillon basées sur des règles séquentielles d'arrêt pourrait être menée de façon similaire. Ainsi, pour l'estimateur de Breslow $\hat{\Lambda}_n$ de Λ_0 défini (voir par exemple [AG82]) pour $t \in [0, \tau]$ par :

$$\hat{\Lambda}_n(t) = \sum_{\{i; X_i \leq t\}} \frac{\delta_i}{\sum_{j \in \mathcal{R}_i} e^{\hat{\beta}_n Z_j(X_i)}},$$

nous pourrions étudier le comportement asymptotique de l'estimateur de la fonction de risque cumulé pour une taille d'échantillon issue d'une procédure séquentielle d'arrêt. En outre, une étude similaire pourrait être menée sur l'estimateur de Kaplan-Meier [KM58], basé sur des bandes de confiance séquentielles (voir par exemple [Gil92]).

Par ailleurs, on peut voir par exemple dans [ABGK93] que de nombreux tests sont basés sur le modèle de Cox, il pourrait être intéressant de considérer la version séquentielle de la statistique de test et de voir dans quelle mesure les résultats de convergence sont prolongés.

Ces résultats ouvrent ainsi la porte à de nombreuses perspectives pour d'autres modèles basés sur d'autres estimateurs. Remarquons enfin ici que notre étude implique des résultats plus forts que ceux jusqu'ici obtenus dans le cas du modèle à risques proportionnels de Cox, puisqu'ils impliquent la convergence complète là où initialement seule la convergence en probabilité avait été obtenue dans [AG82].

À notre connaissance, le modèle traité dans [MU90] constitue l'un des modèles les plus généraux pour lesquels des études sur les tests d'homogénéité ont été menées. En effet, il s'agit ici d'un modèle non paramétrique. Notre étude présente certaines restrictions par rapport aux approches de [CJG97, BC01] en ce qui concerne la matrice W , car on suppose les groupes de tailles finies.

Il serait souhaitable dans un premier temps de mener une étude numérique pour le modèle proposé afin d'étudier le comportement de la statistique de test et de vérifier numériquement les résultats obtenus et surtout d'étudier les limites et points forts des techniques utilisées (biais, conditions pour être proche de l'asymptotique, choix de $(\omega_n)_{n \geq 1}$, etc). En outre, il serait également intéressant de mener une étude du comportement asymptotique de la statistique de test dans le cas d'une version plus faible de l'hypothèse (B3), mais aussi d'étudier cette hypothèse pour des données indépendantes et identiquement distribuées.

Dans un second temps, nous pourrions établir des conditions générales sur W permettant de conserver la normalité asymptotique de la statistique de test.

Enfin, le cadre abordé concerne les variables explicatives qui sont indépendantes du temps, il serait intéressant de généraliser nos résultats au cas de variables explicatives fonction du temps, tant pour ce modèle que pour les modèles qui ont déjà fait l'objet d'études comme le modèle de Cox paramétrique ou semi-paramétrique (voir [CJG97, NC99]).

Bibliographie

- [Aal88] O.O. AALEN : Heterogeneity in survival analysis. *Statistics in Medicine*, 7:1121–1137, 1988.
- [Aal94] O.O. AALEN : Effects of frailty in survival analysis. *Statistical Methods in Medical Research*, 3:227–243, 1994.
- [ABGK93] P.K. ANDERSEN, Ø. BORGAN, R.D. GILL et N. KEIDING : *Statistical Models Based on Counting Processes*. Springer-Verlag, 1993.
- [AC96] A. ALIOUM et D. COMMENGES : A proportional hazards model for arbitrarily censored and truncated data. *Biometrics*, 52:512–524, 1996.
- [AG82] P.K. ANDERSEN et R. D. GILL : Cox’s regression model for counting processes : a large sample study. *The Annals of Statistics*, 10(4):1100–1120, 1982.
- [Ans50] F.J. ANSCOMBE : Large-sample theory of sequential estimation. *Biometrika*, 36:455–458, 1950.
- [Ans52] F.J. ANSCOMBE : Large-sample theory of sequential estimation. *Proc. Cambridge Philos. Soc.*, 48:600–607, 1952.
- [Ans53] F.J. ANSCOMBE : Sequential estimation. *Journal of the Royal Statistical Society, Série B*, 15:1–29, 1953.
- [Ans54] F.J. ANSCOMBE : Fixed-sample-size analysis of sequential observations. *Biometrics*, 10:89–100, 1954.
- [Bar46] G.A. BARNARD : Sequential tests in industrial statistics. *Journal of the Royal Statistical Society*, 8, Suppl.:1–21, 1946.
- [BB03] L. BORDES et C. BREUILS : Sequential estimation for the semiparametric proportional hazards model. Rapport de recherche n°2003.01, Avril 2003.
- [BC01] L. BORDES et D. COMMENGES : Asymptotics for homogeneity tests based on a multivariate random effects proportional hazards model. *J. Multivariate Anal.*, 78(1):83–102, 2001.
- [Ber81] R. BERAN : Nonparametric regression with randomly censored survival data. Rapport technique, Dept. Statist., Univ. California, Berkeley, 1981.
- [BGY97] Y. BILIAS, M. GU et Z. YING : Towards a general asymptotic theory for Cox model with staggered entry. *The Annals of Statistics*, 25(2):662–682, 1997.

- [BHC88] J. BRETAGNOLLE et C. HUBER-CAROL : Effects of omitting covariates in cox's model for survival data. *Scand. J. Stat., Theory Appl.*, 15(2):125–138, 1988.
- [BM73] P.K. BATTACHARYA et A. MALLIK : Asymptotic normality of the stopping times of some sequential procedures. *The Annals of Statistics*, 1(6):1203–1211, 1973.
- [CA95] D. COMMENGES et P.K. ANDERSEN : Score test of homogeneity for survival data. *Lifetime Data Analysis*, 1(2):145–159, 1995.
- [CC85] D. CLAYTON et J. CUZICK : Multivariate generalizations of the proportional hazards model. *Journal of The Royal Statistical Society Series A (Statistics In Society)*, 148:82–117, 1985.
- [Cha96] Y.-C. I. CHANG : Sequential fixed size confidence regions for regression parameters in generalized linear model. *Statistica Sinica*, 6:899–916, 1996.
- [Cha01] Y.-C. I. CHANG : Sequential confidence regions of generalized linear models with adaptive designs. *Journal of Statistical Planning and Inference*, 93:277–293, 2001.
- [CJG97] D. COMMENGES et H. JACQMIN-GADDA : Generalized score test of homogeneity based on correlated random effects models. *J. R. Statist. Soc B*, 59(1):157–171, 1997.
- [Cox72] D.R. COX : Regression models and life tables (with discussion). *Journal of the Royal Statistical Society*, B(34):187–220, 1972.
- [Cox75] D.R. COX : Partial likelihood. *Biometrika*, 62(2):269–276, 1975.
- [Cox83] D.R. COX : Some remarks on overdispersion. *Biometrika*, 70:269–274, 1983.
- [CR65] Y.S. CHOW et H. ROBBINS : On the asymptotic theory of fixed-width sequential confidence intervals for the mean. *Annals of Mathematical Statistics*, 36:457–462, 1965.
- [Dab97] D. M. DABROWSKA : Smoothed cox regression. *Ann. Stat.*, 25(4):1510–1540, 1997.
- [Dea92] C. DEAN : Testing for overdispersion in poisson and binomial regression model. *Journal of the American Statistical Association*, 87:451–457, 1992.
- [DG00] A. DMITRIENKO et Z. GOVINDARAJULU : Sequential confidence regions for maximum likelihood estimates. *The Annals of Statistics*, 28(5):p. 1472–1501, 2000.
- [DL89] C. DEAN et J.F. LAWLESS : Tests for detecting overdispersion in poisson regression models. *Journal of the American Statistical Association*, 84:467–472, 1989.
- [Döh00] S. DÖHLER : Consistent hazard regression estimation by sieved maximum likelihood estimators. Preprint [00-16], Universitat Freiburg, 2000.

- [DR29] H.F. DODGE et H.G. ROMIG : A method of sampling inspection. *Bell Syst. Tech. J.*, 8:613–631, 1929.
- [DSW00] S. DATTA, G.A. SATTEN et J.M. WILLIAMSON : Consistency and asymptotic normality of estimators in proportional hazards model with interval censoring and left truncation. *Annals of the institute of Statistical Mathematics*, 52(1):160–172, 2000.
- [Dur97a] G. DURRIEU : *Contribution statistique à l'étude des maladies d'origine mitochondriales*. Thèse de doctorat, Université de Bordeaux I, 1997.
- [Dur97b] G. DURRIEU : Procédures séquentielles non paramétriques de l'estimation du coefficient de régression dans un modèle linéaire. Dans *XXII Ecole d'été de probabilité de Saint-Flour*, 1997.
- [FH90] T.R. FLEMING et D.P. HARRINGTON : *Counting Processes and Survival Analysis*. John Wiley & Sons, Inc, 1990.
- [Gai00] F. GAÜZÈRE : *Approche non paramétrique pour un modèle 3 états avec censures par intervalles - Application à la dépendance*. Thèse de doctorat, Université de Bordeaux 2, Janvier 2000.
- [GD02] Z. GOVINDARAJULU et A. DMITRIENKO : Sequential estimation of the variance of an unknown distribution. *Journal of Statistical Planning and Inference*, 100:221–237, 2002.
- [Gho91] B.K. GHOSH : *A brief history of sequential analysis in Handbook of sequential analysis (Ghosh, B.K.(ed.); Sen, P.K.(ed.))*, pages 1–19. Dekker, New York, 1991.
- [Gil80] R. D. GILL : *Censoring and stochastic integrals*. Mathematical Centre Tracts 124. Amsterdam : Mathematisch Centrum. V, 1980.
- [Gil92] R. D. GILL : Lectures on survival analysis. Dans *XXII Ecole d'été de probabilité de Saint-Flour*, numéro 2, 1992.
- [Gle65] L.J. GLEESER : On the asymptotic theory of fixed-sized sequential confidence bounds for linear regression parameters. *Annals of Mathematical Statistics*, 36:463–467, 1965.
- [GMS97] M. GHOSH, N. MUKHOPADHYAY et P.K. SEN : *Sequential Estimation*. John Wiley and Sons, Inc, 1997.
- [Gov86] Z. GOVINDARAJULU : Recent developments in sequential analysis. ii : Estimation of parameters. *Math. Sci.*, 11:17–31, 1986.
- [Gra95] R.J. GRAY : Tests for variation over groups in survival data. *Journal of the American Statistical Association*, 90(429):198–203, 1995.
- [Gre26] M. GREENWOOD : The natural duration of cancer. *Reports on public health and medical subjects*, 33:1–26, 1926.
- [GSB80] M.H. GAIL, T.J. SANTNER et C.C. BROWN : An analysis of comparative carcinogenesis experiments based on multiple times to tumor. *Biometrics*, 36:255–266, 1980.

- [GW91] P.E. GREENWOOD et W. WEFELMEYER : Efficient estimation in a non-linear counting-process regression model. *Can. J. Stat.*, 19(2):165–178, 1991.
- [Hal45] J.B.S. HALDANE : On a method of estimating frequencies. *Biometrika*, 33:222–225, 1945.
- [HJ94] J. HOFFMANN-JØRGENSEN : *Probability with a view toward statistics*. Chapman & Hall, 1994.
- [HR69] P.L. HSU et H. ROBBINS : Complete convergence and the law of large numbers. *Proc. Nat. Acad. Sci. U.S.A.*, 33:25–31, 1969.
- [Hub81] P.J. HUBER : *Robust statistics*. Wiley Series in Probability and Mathematical Statistics, 1981.
- [Hub00] C. HUBER : *Recent advances in reliability theory. Methodology, practice and inference (Limnios, N.; Nikulin, M. (editors))*, chapitre 19. Birkhäuser, 2000.
- [JGC95] H. JACQMIN-GADDA et D. COMMENGES : Tests of homogeneity for generalized linear models. *Journal of the American Statistical Association*, 90(432):1237–1246, 1995.
- [Kle92] J.P. KLEIN : Semiparametric estimation of random effects using the cox model based on the em algorithm. *Biometrics*, 48:749–806, 1992.
- [KLR91] J.D. KALBFLEISCH, J.F. LAWLESS et J.A. ROBINSON : Methods for the analysis and prediction of warranty claims. *Technometrics*, 33(3):273–285, 1991.
- [KM58] E.L. KAPLAN et P. MEIER : Nonparametric estimation from incomplete observations. *Journal of the American Statistical Association*, 53:457–481, 1958.
- [KM97] J.P. KLEIN et M.L. MOESCHBERGER : *Survival Analysis, Techniques for Censored and Truncated Data*. Springer, Statistics for Biology and Health, 1997.
- [KP80] J.D. KALBFLEISCH et R.L. PRENTICE : *The statistical analysis of failure time data*. Wiley Series in Probability and Mathematical Statistics. New York etc. : John Wiley and Sons, 1980.
- [Lai01] T. L. LAI : Sequential analysis : some classical problems and new challenges (with discussion). *Statistica Sinica*, 11:303–408, 2001.
- [Lan90] T. LANCASTER : *The econometric analysis of transition data*. Econometric Society Monographs, 17. Cambridge etc. : Cambridge University Press, 1990.
- [Let00] F. LETUÉ : *Modèle de Cox : estimation par sélection de modèle et modèle de chocs bivarié*. Thèse de doctorat, Université de Paris Sud - No6414, décembre 2000.
- [Lia87] K.-Y. LIANG : A locally most powerful test for homogeneity with many strata. *Biometrika*, 74:259–264, 1987.

- [LN95] J.F. LAWLESS et C. NADEAU : Some simple robust methods for the analysis of recurrent events. *Technometrics*, 37(2):158–168, 1995.
- [LY94] D.Y. LIN et Z. YING : Semiparametric analysis of the additive risk model. *Biometrika*, 81(1):61–71, 1994.
- [LY95a] D.Y. LIN et Z. YING : Semiparametric analysis of general additive-multiplicative hazard models for counting processes. *The Annals of Statistics*, 23(5):1712–1734, 1995.
- [LY95b] D.Y. LIN et Z. YING : Semiparametric inference for the accelerated life model with time- dependent covariates. *J. Stat. Plann. Inference*, 44(1):47–63, 1995.
- [MD97] W. MUKHOPADHYAY et W.T. DUGGAN : Can a two-stage procedure enjoy second-order property? *The Indian Journal of Statistics*, 59(3 A):435–448, 1997.
- [MS98] I.W. MCKEAGUE et S. SUBRAMANIAN : Product-limit estimators and cox regression with missing censoring information. *Scandinavian Journal of Statistics*, 25(4):589–601, 1998.
- [MS02] T. MARTINUSSEN et T.H. SCHEIKE : A flexible additive multiplicative hazard model. *Biometrika*, 89(16):283–296, 2002.
- [MU90] I.W. MCKEAGUE et K.J. UTIKAL : Inference for a nonlinear counting process regression model. *The Annals of Statistics*, 18(3):1172–1187, 1990.
- [MvdV00] S.A. MURPHY et A.W. van der VAART : On profile likelihood. *Journal of the American Statistical Association*, 95(450):449–485, 2000.
- [Nae82] T. NAES : The asymptotic distribution of the estimator for the regression parameter in cox’s regression model. *Scandinavian Journal of Statistics*, 9:107–115, 1982.
- [NC97] E.T.M. NG et R.J. COOK : Modeling of two-state disease processes with random effects. *Lifetime Data Analysis*, 3(4):315–335, 1997.
- [NC99] E. T.M. NG et R. J. COOK : Adjusted score tests of homogeneity for poisson processes. *J. Am. Stat. Assoc.*, 94(445):308–319, 1999.
- [NGAS92] G.G. NIELSEN, R.D. GILL, P.K. ANDERSEN et T.I.A. SØRENSEN : A counting process approach to maximum likelihood estimation in frailty models. *Scandinavian Journal of Statistics*, 19(1):25–43, 1992.
- [Pol90] D. POLLARD : *Empirical Processes : Theory and Applications*. NSF-CBMS Regional Conference Series in Probability and Statistics, Volume 2, 1990.
- [She31] W.A. SHEWHART : *Economic control of manufactured products*. Van Nostrand Reinhold, New York, 1931.
- [Sie03] D. SIEGMUND : Invited paper : Herbert robbins and sequential analysis. *The Annals of Statistics*, 31(2):349–365, 2003.

- [SK86] C.A. STRUTHERS et J.D. KALBFLEISCH : Misspecified proportional hazard models. *Biometrika*, 73:363–369, 1986.
- [SS83] T. SELLKE et D. SIEGMUND : Sequential analysis of the proportional hazards model. *Biometrika*, 70(2):315–326, 1983.
- [Ste45] C. STEIN : A two-sample test for a linear hypothesis whose power is independent of the variance. *Annal of mathematical Statistics*, 16:243–258, 1945.
- [Ste49] C. STEIN : Some problems in sequential estimation. *Econometrica*, 17:77–78, 1949.
- [SW76] R.J. SERFLING et D.D. WACKERLY : Asymptotic theory of sequential fixed-width confidence interval. *Journal of the American Statistical Association*, 71(356):949–955, 1976.
- [SW86] G.R. SHORACK et J.A. WELLNER : *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics, 1986.
- [Tho33] W.R. THOMPSON : On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25:285–294, 1933.
- [Tsi81] A.A. TSIATIS : A large sample study of Cox’s regression model. *The Annals of Statistics*, 9(1):93–108, 1981.
- [vdV02] A. van der VAART : Semiparametric statistics. Dans SPRINGER, éditeur : *Lectures on probability theory and statistics.*, Lect. Notes Math. 1781, pages 331–457. Ecole d’été de probabilités de Saint-Flour XXIX - 1999, Saint-Flour, France, July 8-24, 1999 2002.
- [vdVW96] A. van der VAART et J.A. WELLNER : *Weak convergence and empirical processes*. Springer, 1996.
- [Wal46] A. WALD : Some improvements in setting limits for the expected number of observations required by a sequential probability ratio test. *Annal of mathematical Statistics*, 17:466–474, 1946.
- [Wei51] W. WEIBULL : A statistical distribution function of wide applicability. *J. appl. Mech.*, 18:293–297, 1951.
- [Won86] W.H. WONG : Theory of partial likelihood. *The Annals of Statistics*, 14(1):88–123, 1986.
- [WW48] A. WALD et J. WOLFOWITZ : Optimum character of the sequential probability ratio test. *Annal of mathematical Statistics*, 19:326–339, 1948.

MOTS-CLÉS : méthodes séquentielles, consistance asymptotique, efficacité asymptotique, modèle à risques proportionnels de Cox, inégalités exponentielles.

RÉSUMÉ. Ce mémoire porte sur deux sujets distincts dans le cadre de l'analyse statistique de durées de vie. Il s'articule donc en deux parties.

La première partie concerne l'estimation séquentielle du paramètre de régression pour le modèle à risques proportionnels de Cox dans le cadre de données censurées à droite. Dans un premier temps, on montre comment il est possible de définir des règles d'arrêt permettant de garantir une certaine qualité de l'estimation. Ces règles d'arrêt conduisent alors à des estimateurs qui dépendent de tailles d'échantillons aléatoires pour lesquels on démontre que le comportement asymptotique (au sens d'un degré de précision croissant) est le même que celui des estimateurs non séquentiels. Les propriétés démontrées sont ensuite étendues au cadre multidimensionnel et des simulations de Monte-Carlo illustrent les comportements des estimateurs séquentiels. Enfin cette première partie s'achève par l'étude théorique du comportement de la variable d'arrêt dans le cadre d'intervalles de confiance séquentiels. Il est établi que la règle d'arrêt, correctement normalisée, est asymptotiquement normale.

La seconde partie de ce mémoire porte sur la construction de tests d'homogénéité dans le cadre d'un modèle de durées de vie non paramétrique incluant des variables explicatives ainsi qu'un phénomène de censure à droite. Une statistique de test est proposée et son comportement asymptotique est établi. Ces résultats viennent compléter des résultats récents sur les tests d'homogénéité basés sur une statistique de *score*. Cependant la complexité du modèle rend l'étude théorique de la statistique de test plus difficile que celle des modèles paramétriques ou semi-paramétriques qui ont été étudiés jusque là. En particulier, les résultats asymptotiques obtenus ne sont utilisables que pour tester des effets de groupes.

KEYWORDS : sequential methods, asymptotic consistency, asymptotic efficiency, proportional hazards model, exponential inequalities

ABSTRACT. This thesis deals with two distinct subjects around lifetime data statistical analysis and, thus, it is divided into two parts.

The first one deals with sequential estimation for the Cox proportional hazards model in the case of right-censored data. First we describe the way to define stopping rule with a certain quality of estimation. It leads us to consider random sample sizes and we show that the asymptotic behaviour of the sequential estimations is similar with the deterministic ones (asymptotic takes the sense where the precision degree is increasing). We need to use various mathematical tools such as exponential inequalities for empirical processes in addition to some techniques inspired by the martingale theory. We then extend these properties for the multidimensional case and Monte-Carlo simulations give us an numerical approach for the asymptotic behaviour of sequential estimates. This first part ends with a theoretical study of the stopping variable behaviour in the frame of sequential confidence intervals. We establish there the asymptotic normality of the stopping rule under convenient normalisation.

Our aim in the second part is to build homogeneity tests in the frame of nonparametric lifetime data model including right-censorship and covariates. A test is established and its asymptotic law is obtained. This results allow us to complete recent properties of statistics tests based on score function. The theoretical result requires complex technical results and lead us to show the asymptotic results only to test groups effects.